

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:11:16 ; Search time 825.255 Seconds
(without alignments)
1233.024 Million cell updates/sec

Title: US-10-777-131A-1

Perfect score: 21

Sequence: 1 tccggatgatgatcttaccctc 21

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ha: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sbs: *
12: gb_gy: *
13: gb_un: *
14: gb_vl: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	6	AX770103 Sequence
2	21	100.0	823	6	AX770121 Sequence
3	21	100.0	3002	6	CQ880370 Sequence
4	21	100.0	40056	1	MSGY348
5	21	100.0	110000	1	AE000516_12
6	21	100.0	327650	1	BX248337 Mycobacte
7	21	100.0	349306	1	BX248337 Mycobacte
8	17.8	84.8	144190	2	AP004112 Oryza sat
9	17.8	84.8	166453	8	AP005649 Oryza sat
10	17.8	84.8	175567	9	AC079447 Homo sapi
11	17.8	84.8	195744	2	AC092002 Homo sapi
12	17.8	84.8	247746	2	AC112285 Rattus no
13	17.8	84.8	291544	2	AC128496 Rattus no
14	17.8	84.8	320731	2	AC149573 Medicago
15	17.4	82.9	1286	8	ATU90428 Arabidopsi
16	17.4	82.9	109375	8	AC003672 Arabidops
17	17.4	82.9	152709	2	AC149508 Strongylo
18	17	81.0	472	10	BC025166 Mus muscu
19	17	81.0	4578	10	AF104410 Mus muscu

C	20	17	81.0	11013	1	AE011443	AE011443 Leptospir
C	21	17	81.0	109431	2	AC016696	AC016696 Homo sapi
C	22	17	81.0	138905	2	AC131507	AC131507 Strongylo
C	23	17	81.0	144228	9	AC068599	AC068599 Homo sapi
C	24	17	81.0	223276	10	AL593853	AL593853 Mouse DNA
C	25	17	81.0	300478	1	AE017201	AE017201 Lactobaci
C	26	17	81.0	301508	1	AE017292	AE017292 Leptospir
C	27	17	81.0	349980	6	AY926711	AY926711 Sequence
C	28	17	81.0	349980	6	AX926712	AX926712 Sequence
C	29	16.8	80.0	519	3	AY488869	AY488869 Plasmodu
C	30	16.8	80.0	519	3	AY488871	AY488871 Plasmodu
C	31	16.8	80.0	4308	6	AR528296	AR528296 Sequence
C	32	16.8	80.0	10358	1	AE013482	AE013482 Melnansa
C	33	16.8	80.0	28149	8	YSC19931	Y00162 Saccharomyc
C	34	16.8	80.0	90132	8	AC084165	AC084165 Arabidops
C	35	16.8	80.0	110000	1	CR522870_29	Continuation (30 o
C	36	16.8	80.0	117919	8	AP003220	AP003220 Oryza sat
C	37	16.8	80.0	131239	9	HS667H12	AL035414 Human DNA
C	38	16.8	80.0	133060	2	AC138555	AC138555 Carollia
C	39	16.8	80.0	137678	9	HS879X22	AL034351 Human DNA
C	40	16.8	80.0	164138	2	AC109432	AC109432 Rattus no
C	41	16.8	80.0	166143	2	AL356960	AL356960 Homo sapi
C	42	16.8	80.0	169486	8	AP005648	AP005648 Oryza sat
C	43	16.8	80.0	208155	9	AC124254	AC124254 Homo sapi
C	44	16.8	80.0	228736	2	AC107161	AC107161 Rattus no
C	45	16.4	78.1	3292	1	AY038363	AY038363 Shewanell

ALIGNMENTS

RESULT 1	AX770103	Sequence 1 from Patent WO03016562.	21 bp	DNA	linear	PAT 02-JUL-2003
LOCUS	AX770103					
DEFINITION	AX770103					
ACCESSION	AX770103					
VERSION	AX770103.1	GI:32437681				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						
ORIGIN						
Query Match						
Best Local Similarity						
Matches						
Qy						
Db						
RESULT 2	AX770121	Sequence 19 from Patent WO03016562.	823 bp	DNA	linear	PAT 02-JUL-2003
LOCUS	AX770121					
DEFINITION	AX770121					
ACCESSION	AX770121					
VERSION	AX770121.1	GI:32437699				
KEYWORDS						
SOURCE						
ORGANISM						

REFERENCE 1 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

AUTHORS 1 Glacquel, B.

JOURNAL 1 Compositions and methods for detecting multidrug resistant strains of M. tuberculosis having mutations in genes of the mutR family

FEATURES 1 Patent: WO 03016562-A 19 27-FEB-2003; INSTITUT PASTEUR (FR)

SOURCE 1 Location/Qualifiers

ORIGIN 1 .823 /organism="Mycobacterium tuberculosis" /mol_type="unassigned DNA" /db_xref="taxon:1773"

Query Match 100.0%; Score 21; DB 6; Length 823; Best Local Similarity 100.0%; Pred. No. 1.9; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCGATGATGATTACCTCC 21 5 TCCGATGATGATTACCTCC 25

RESULT 3 3002 bp DNA linear PAT 11-OCT-2004

LOCUS CO880370

DEFINITION Sequence 1 from Patent WO2004083459.

ACCESSION CO880370

VERSION CO880370.1 GI:54034136

KEYWORDS

SOURCE Mycobacterium tuberculosis

ORGANISM Mycobacterium tuberculosis Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1 Bange, F.C. Method and kit for a specific detection of m. tuberculosis

AUTHORS 1 Patent: WO 2004083459-A 1 30-SEP-2004; Artus Gesellschaft Fuer Molekularbiologische Diagnostik und Entwicklung MBH. (DE)

JOURNAL 1 Location/Qualifiers

FEATURES 1 .3002 /organism="Mycobacterium tuberculosis" /mol_type="unassigned DNA" /db_xref="taxon:1773"

SOURCE

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 3002; Best Local Similarity 100.0%; Pred. No. 2.3; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCGATGATGATTACCTCC 21 788 TCCGATGATGATTACCTCC 808

RESULT 4 40056 bp DNA linear BCT 10-DEC-1996

LOCUS MSGY348

DEFINITION Mycobacterium tuberculosis sequence from clone y348.

ACCESSION ADO000020

VERSION ADO000020.1 GI:1717739

KEYWORDS

SOURCE Mycobacterium tuberculosis

ORGANISM Mycobacterium tuberculosis Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1 Du, L. (bases 1 to 40056)

TITLE Direct Submission

JOURNAL Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA, USA, 02154 du@eric.com

COMMENT GSDB:8:1004716.

FEATURES 1 Location/Qualifiers

SOURCE 1 .40056 /organism="Mycobacterium tuberculosis" /mol_type="genomic DNA" /db_xref="taxon:1773" /clone="y348"

ORIGIN

Query Match 100.0%; Score 21; DB 1; Length 40056; Best Local Similarity 100.0%; Pred. No. 3.5; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCGATGATGATTACCTCC 21 17919 TCCGATGATGATTACCTCC 17939

RESULT 5

AB000516.12

WPCOMMENT

Sequence split into 44 fragments LOCUS AB000516 Accession AB000516

Fragment Name	Begin	End
AB000516_00	1	110000
AB000516_01	100001	210000
AB000516_02	200001	310000
AB000516_03	300001	410000
AB000516_04	400001	510000
AB000516_05	500001	610000
AB000516_06	600001	710000
AB000516_07	700001	810000
AB000516_08	800001	910000
AB000516_09	900001	1010000
AB000516_10	1000001	1110000
AB000516_11	1100001	1210000
AB000516_12	1200001	1310000
AB000516_13	1300001	1410000
AB000516_14	1400001	1510000
AB000516_15	1500001	1610000
AB000516_16	1600001	1710000
AB000516_17	1700001	1810000
AB000516_18	1800001	1910000
AB000516_19	1900001	2010000
AB000516_20	2000001	2110000
AB000516_21	2100001	2210000
AB000516_22	2200001	2310000
AB000516_23	2300001	2410000
AB000516_24	2400001	2510000
AB000516_25	2500001	2610000
AB000516_26	2600001	2710000
AB000516_27	2700001	2810000
AB000516_28	2800001	2910000
AB000516_29	2900001	3010000
AB000516_30	3000001	3110000
AB000516_31	3100001	3210000
AB000516_32	3200001	3310000
AB000516_33	3300001	3410000
AB000516_34	3400001	3510000
AB000516_35	3500001	3610000
AB000516_36	3600001	3710000
AB000516_37	3700001	3810000
AB000516_38	3800001	3910000
AB000516_39	3900001	4010000
AB000516_40	4000001	4110000
AB000516_41	4100001	4210000
AB000516_42	4200001	4310000
AB000516_43	4300001	4403837

Continuation (13 of 44) of AB000516 from base 1200001 (AB000516 Mycobacterium tuberculosis

Query Match 100.0%; Score 21; DB 1; Length 110000;

Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TCCGATGATGATTTACTCTCC 21
|||||
db 85887 TCCGATGATGATTTACTCTCC 85907
|||||

RESULT 6
BX248337 327650 bp DNA linear BCT 06-MAY-2004
LOCUS 327650 bp
DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment 4/14.
ACCESSION BX248337 BX248337
VERSION BX248337.1 GI:31617663
KEYWORDS complete genome.
SOURCE Mycobacterium bovis AF2122/97
ORGANISM Mycobacterium bovis AF2122/97
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE 1
AUTHORS Garnier,T., Biglmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
Pryor,M., Duthoy,S., Grondin,S., Lacroix,C., Monsemp,C., Simon,S.,
Harris,B., Aikin,R., Doggett,J., Mayes,R., Keating,L.,
Wheeler,P.R., Parkhill,J., Barrell,B.G., Cole,S.T., Gordon,S.V. and
Hewison,G.
The complete genome sequence of Mycobacterium bovis
Unpublished
2 (bases 1 to 327650)
Garnier,T.
Direct Submission

Submitted (24-MAR-2003) Garnier T., Unite de Genetique Molculaire
Bacterieme Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK. PTH Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Molculaire Bacterieme, Institut Pasteur, 28
rue du Docteur Roux, 75724 Paris Cedex 15, France

FEATURES
SOURCE location/Qualifiers
1. 327650

gene /organism="Mycobacterium bovis AF2122/97"
/mol_type="genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"
698..2170
/locus_tag="Mb0912"
698..2170
/locus_tag="Mb0912"
/note="Mb0912", len: 490 aa. Equivalent to Rv0888, len:
490 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 490 aa overlap). Probable exported
protein. Equivalent to AAK45157.1 from Mycobacterium
tuberculosis strain CDC1531 (507 aa) but shorter 17 aa.
Contains possible N-terminal signal sequence."

cds /codon_start=1
/transl_table=1
/product="PROBABLE EXPORTED PROTEIN"
/protein_id="CAD93773.1"
/db_xref="GI:31617664"
/db_xref="GOA:Q10549"
/db_xref="UniProt/Swiss-Prot:Q10549"
/translation="MDYAKRIGVGALAVVGVGAAYTTAIGSAAPTPSSSTDSBP
VDACSPGGSASLSLAIFGASVPOVQAVDPCGIPDILNALIDFLAAVANGVPII
ENRTVPANPGGASVGGVGVGFADCDPGRNMTPAVERGA PGGQHGIVVDR
TASRTADBSFGVTDPSPVSDTSHVHGAGVAGCPFGHGDVAATVTVFVNTPT
DTISGDSMLTYNAGLPPLSSALIRFFTKETGKLNAYVYANVOEDPAHYQFLI
KSKMPSOTPEPEPTLLMPICGVPSDGLNTLSERKVOQLDQTVTECTSDCLTKGP
TYSQMRLPGGTVVYNLHTVTGGPTTNALQAVANYIQNSAGRAIVTVGDNNARY
SDDSALIQFAOVNGLTDVAWQVEHGPFTTPEPAFCMVGNCEBLDKIFYSGGVTL

gene
cds

GAVSNGNAPKRFNSKGEPLSDHSFPAVGVHYADVNAVNR"
complement(12205..3326)
/gene="cItA"
/locus_tag="Mb0913c"
complement(12205..3326)
/gene="cItA"
/locus_tag="Mb0913c"
/EC_number="2.3.3.1"
/note="Mb0913c, cItA, len: 373 aa. Equivalent to Rv0889c,
len: 373 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 373 aa overlap). Probable cItA
alternate gene name: gItA, citrate synthase 2 (EC
4.1.3.7), highly similar to others e.g.
CAB58899.1|A45998 putative citrate synthase from
Streptomyces coelicolor (387 aa); P39119|CISY_BACSU
citrate synthase II from Bacillus subtilis (366 aa), FASTA
scores: opt: 586, E(): 5.8e-30, (33.8% identity in 367 aa
overlap); etc. Also similar to Rv0896|MYCY31.24 from
Mycobacterium tuberculosis (29.2% identity in 274 aa
overlap) and Rv1131. Contains P500480 Citrate synthase
signature. BELONGS TO THE CITRATE SYNTHASE FAMILY."

cds /codon_start=1
/transl_table=1
/product="PROBABLE CITRATE SYNTHASE II CITA"
/protein_id="CAD93774.1"
/db_xref="GI:31617665"
/db_xref="GOA:Q10529"
/db_xref="UniProt/Swiss-Prot:Q10529"
/translation="MTVPENFVGLDGVAAFTTEIEAPDKGALRYGVDIRDLVS
QRTVFGVMAVLVDGNGSLPAPPEPLPIHSGDVVDVQAGLAMPITWGVPLD
IDATARQOLARASVMAISYVAOSARGIYOPAVORTIDECTAFRMTWGSEDP
RHTEAIDAVVSAASEHGMNSTPAPRATYGTADVAALSGAIGANGPRTGAPRY
LPMIDVERAGDARSVYKGLIDRCEKLMGSHRYRABDPRAVRAAEYLGAPRYE
VAAVEQAALSEKRRPDAIEFNVEFAVVDFAVVPANMPAMFTCGRTGWCA
HIIQKRIKGLVPSAISYVGPSPESVDGWERVLTGA"
complement(13413..6061)
/locus_tag="Mb0914c"
complement(13413..6061)
/locus_tag="Mb0914c"
/note="Mb0914c", len: 882 aa. Equivalent to Rv0890c,
len: 882 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.9% identity in 882 aa overlap). Probable
transcriptional regulatory protein, luxR family, highly
similar (but shorter 238 aa in N-terminus) to
NP_302202.1|NC_002677 possible transcriptional regulator
from Mycobacterium leprae (1106 aa). Also highly similar
(generally in part) to others e.g. T50568 probable
multi-domain regulatory protein from Streptomyces
coelicolor (1334 aa); P10957|NARL_ECOLI nitrate/nitrite
response regulator protein from Escherichia coli (216 aa),
FASTA scores: opt: 193, E(): 6e-06, (37.4% identity in 93
aa overlap); etc. Also highly similar to others from
Mycobacterium tuberculosis e.g. MTCY02810.22, MTVO08.44,
MTY036.21, and MTCY31.24. Contains P500017 ATP/GTP-binding
site motif A (P-loop), P500622 Bacterial regulatory
proteins, luxR family signature, and probable helix-turn
helix motif from aa 836 to 857 (score 1559, +4.50 SD).
BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
REGULATORS."

cds /codon_start=1
/transl_table=1
/product="PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN
(PROBABLY LUXR-FAMILY)"
/protein_id="CAD93775.1"
/db_xref="GI:31617666"
/db_xref="UniProt/Swiss-Prot:P59969"
/translation="MRALLQNLRLVLTGCGVGTTLAQLIASASLEBDGLCFVDLA
PIRESGLVATARAVALPDPQGRSTDSLRFRIGNRNMLVDNCHLIDACAAVY
ELGACPELTILASREPLIMAGEITRMVSPETIDPAVEFLDRAAGVAGPFIYAH
NAAVGEICRLDGIPLAIFPAARVRSMSPLEIADLDCFLDRAAGVAGVROOT
LRASIDSHALLTETEDILFRRLAPFVGGFDLAVRVAAGSDIDPVSVDQLTLVD
KSLVAVDDCGRRRLVETVRVYALEKDGSEADVAHARRDYTTALASLNPAN
DHQRLVARAETEDINLAAFAWSERENCHITAEALQLASSLOPIWFGRAHLREGLSWFS

probable short-chain dehydrogenase from *Pseudomonas aeruginosa* (253 aa): AAC44307.1|US9433.3-keoacyl-acyl carrier protein reductase from *Bacillus subtilis* (246 aa); etc. Also similar to other dehydrogenases from *Mycobacterium tuberculosis* e.g. MTCY09F9.36, E(1):1.4e-18; MTCY369.14, E(1):8e-17; MTCY02B10.14, E(1):2.5e-14; MTCY09F9.23c, E(1):1.5e-13; MTCY03C7.07, E(1):1.9e-13. Contains PS00061 Short-chain dehydrogenases/reductases family signature, and PS00017 ATP/GTP-binding site motif A (P-loop). BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY."

/transl_table=1
/codon_start=1
/product="PROBABLE SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE"
/protein_id="CAB08497.1"
/db_xref="GI:2078076"
/translation="MILDMFRLDKRAVITGGGRLGAALAPAAQAGADVLASRTS SELDAVAEQIAPAGRAHTVAADLAHPVLAAGAVAGFKLDIVYNNQGTMP LTSTDLADAPAPNGVGTALTAAPVAPLMLHSGSGSVINISITWRLARGPAAV GTAKALAHYTRIALADLCPRVRNALTPSGSILTSALFVAVANDELRAPMEQATPLRR LGDPVDIAAAVYASPSGSPFTGKTLFVSDGLTFPNLDLPDL"

/complement(113..136)
/locus_tag="Rv0927c"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
/complement(320..406)
/locus_tag="Rv0927c"
/note="PS00061 Short-chain alcohol dehydrogenase family signature"
1116..2228
/gene="pbcS3"
/locus_tag="Rv0928"
1116..2228
/gene="pbcS3"
/locus_tag="Rv0928"
/function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT). THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE TRANSPORT."

/standard_name="phoS2"
/note="Rv0928, (MTCY21C12.22), len: 370 aa. pbcS3 (previously known as phoS2), phosphate-binding lipoprotein component of inorganic phosphate transport system (see citations below), highly similar to others from *Mycobacterium leprae* e.g. Q50089|PST83|PHOS1 phosphatase-binding protein 3 precursor (328 aa), FASTA scores: opt: 1772, E(1): 0, (79.6% identity in 328 aa overlap); and highly similar to others e.g. AAF74819.1|AF137360.1|AF137360 periplasmic phosphate permease from *Mycobacterium avium* (369 aa). Also highly similar to Rv0932c|MTCY08D9.07|pbcS2 PHOSPHATE-BINDING PERIPLASMIC LIPOPROTEIN (370 aa); and Rv0934|pbcS1 PHOSPHATE-BINDING PERIPLASMIC LIPOPROTEIN (374 aa) from *Mycobacterium tuberculosis* (*Mycobacterium tuberculosis* seems to have three PbcS-like proteins, others being Rv0932c and Rv0934c). Contains lipoprotein signature (PS00013) at N-terminus. BELONGS TO FAMILY OF PHOSPHATE RECEPTORS FOR BACTERIAL ABC-TYPE LIPOPROTEIN TRANSPORTERS."

/codon_start=1
/evidence=experimental
/transl_table=1
/product="PERIPLASMIC PHOSPHATE-BINDING LIPOPROTEIN PSTS3 (PBP-3) (PST83) (PHOS1)"
/protein_id="CAB55337.1"
/db_xref="GI:41353620"
/translation="MKLNRFGAAVGLAALVLASCGNDNVTGATTTGQASAKV DCGKTKLAKSGTAQANMTFRVNFEOACGGTQNTYANGSGAGISEFNQDPEG SDPLSKDEAAARCGSPAMNLPLVPGFATVATYNNLINDGTLTKTNGSTI TQNNPRAIQANBDPTLGERIHVFRDESGTDPNRPRIQAASNGWGAQKGSFO GGVBGARGNDGTSAAAKNTPGSITYNWSTRAQOHLPMANIVSAGDPAITIDSV GQTLAGATISGVGNDLVLDTSFPRPKPGSYPLVLAETVYCSKPSDSQVGTAVKAP LQSTIGAGOSGIGNGYIPIDEPKSRISTLVNMAIA"

misc_feature 1152..1184
/gene="pbcS3"
/locus_tag="Rv0928"
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
2241..3215
/gene="pbcC2"
/locus_tag="Rv0929"
2241..3215
/gene="pbcC2"
/locus_tag="Rv0929"
/function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE. THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE TRANSPORT."
/note="Rv0929, (MTCY21C12.23), len: 324 aa. pbcC2, phosphate-transport integral membrane ABC transporter (see citations below), highly similar to others e.g. NP_302394.1|NC_002677 membrane-bound component of phosphate transport from *Mycobacterium leprae* (319 aa); CAB88474.1|ALJ53816 phosphate ABC transport system permease protein from *Streptomyces coelicolor* (336 aa); NP_290359.1|NC_002655 high-affinity phosphate-specific transport system (cytoplasmic membrane component) from *Escherichia coli* strain 0157:H7 (319 aa); etc. Also similar to Rv935|MTCY08D9.04|PSTC1 PROBABLE TRANSMEMBRANE ABC TRANSPORTER COMPONENT OF PHOSPHATE UPTAKE SYSTEM FROM *Mycobacterium tuberculosis* (338 aa). Contains binding-protein-dependent transport systems inner membrane component signature (PS00402)."
/codon_start=1
/transl_table=1
/product="PHOSPHATE-TRANSPORT INTEGRAL MEMBRANE ABC TRANSPORTER PBC2"
/protein_id="CAB08489.1"
/db_xref="GI:3261752"
/translation="MTVEPLTPKPLVAVDMRPARGRRLFKLASAGSTIVAILLI ALPLVAVPSIRANHANPFTSQFQSDDEQLARGDLFMTALSSITLVAVPV AVGIAVLTHYAPRLRSPPGAMVDLLAAPSITFGMGFVLAPKLEPIRFLNRLL GWLFLRQGVNSLAGGTITPAGIVLSVMLIPVTSISREVFRTPLIQIEAALAGA TKNEVMTVLPGRSQGVVASMIGLRAIGETVAIVVIRASAPGWTSIFDGGYTF ASKIASAASESEPLPGAYISAGFALFVLTVNAAARAAGKVG"
2844..2930
/gene="pbcC2"
/locus_tag="Rv0929"
/note="PS00402 Binding-protein-dependent transport systems inner membrane comp signature"
3212..4138
/gene="pbcA1"
/locus_tag="Rv0930"
3212..4138
/gene="pbcA1"
/locus_tag="Rv0930"
/function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE."

Query Match 100.0%; Score 21; DB 1; Length 349306;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 TCCGATGATGATTACTTC 21
|||||
Db 252612 TCCGATGATGATTACTTC 252632

RESULT 8
AP004112
LOCUS ORyza sativa (japonica cultivar-group) chromosome 2 clone
DEFINITION OJ108 H04, *** SEQUENCING IN PROGRESS ***.
ACCESSION AP004112

VERSION AP004112.1 GI:15375098
 HTG: HTGS PHASE2.
 KEYWORDS Oryza sativa (japonica cultivar-group)
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriactoidae; Oryzaceae; Oryza.

REFERENCE
 1 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, BAC
 clone: OJ108_H04
 Published Only in Database (2001)

JOURNAL
 2 (bases 1 to 144190)
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Direct Submission
 Submitted (29-AUG-2001) Takuji Sasaki, National Institute of
 Agricultural Sciences, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://rsgp.dna.affrc.go.jp/,
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)
 The nucleotide sequence of this BAC clone was generated by
 combining Monsanto and RGP-Japan sequencing data.

COMMENT
 NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1..144190
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="2"
 /clone="OJ108_H04"

ORIGIN
 Query Match 84.8%; Score 17.8; DB 2; Length 144190;
 Best Local Similarity 90.5%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY 1 TCCGATGATGATTAACCTCC 21
 |||||
 Db 73873 TCCGATGATGATTAACCTCC 73893

RESULT 9
 AP005649 166453 bp DNA linear PLN 24-MAR-2004
 LOCUS Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 2,
 DEFINITION BAC clone: OSUNBA0064G16.
 ACCESSION AP005649
 VERSION AP005649.3 GI:45736185
 KEYWORDS Oryza sativa (japonica cultivar-group)
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriactoidae; Oryzaceae; Oryza.

REFERENCE
 1 Sasaki, T., Matsumoto, T. and Katayose, Y.
 Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, BAC
 clone: OSUNBA0064G16
 Published Only in Database (2002)
 JOURNAL
 2 (bases 1 to 166453)
 Sasaki, T., Matsumoto, T. and Katayose, Y.
 Direct Submission
 Submitted (21-AUG-2002) Takuji Sasaki, National Institute of

COMMENT
 Agrobiological Sciences, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://rsgp.dna.affrc.go.jp/,
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)
 On Mar 25, 2004 this sequence version replaced gi:32440965.
 Genes were predicted from the integrated results of the following:
 GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
 (http://www.softberry.com/), GeneMark.hmm
 (http://opal.biology.gatech.edu/GeneMark/), glimmer
 (http://www.tigr.org/tdb/glimmer/glimmer-form.html), RICEHMM
 (http://rsgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
 (http://bioinformatics.laetate.edu/cgi-bin/sp.cgi), SIm4
 (http://glofin.cse.psu.edu/html/docs/sim4.html), gap2
 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The
 genomic sequence was searched against NCBI NonRedundant Protein
 database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
 sequence database at RGP or DDBJ. Protein homologues of the coding
 regions were searched against NCBI NonRedundant Protein database
 with BLASTP. ESTs represent the identified cDNA sequences using
 BLASTN with the corresponding DDBJ accession no. and RGP clone ID.
 Full-length cDNAs represent the identified cDNA sequences using
 BLASTN with the corresponding DDBJ accession no.
 A gene with identity or significant homology to a protein is
 classified based on the protein name to indicate the homology level
 such as same name, 'putative-' and '-like protein'. A gene without
 significant homology to any protein but with full-length cDNA or
 EST homology (covering almost the entire length of partial
 sequence) is classified as an 'unknown' protein. A gene predicted
 by two or more gene prediction programs is classified as a
 'hypothetical' protein according to IRISP standard. A gene
 predicted by a single gene prediction program is also classified as
 a probable 'hypothetical' protein and is included as a
 miscellaneous feature of the sequence.
 The orientation of the sequence is from -21M13 to M13rev of the BAC
 clone. This sequence of OSUNBA0064G16 clone has an overlap with
 OSUNBA0018G11 (DDBJ: AP005776) clone at 5' end and with OJ108_C03
 (DDBJ: AP005288) clone at 3' end. Detailed information on overlap
 and assembly quality together with annotation of this entry is
 available at http://rsgp.dna.affrc.go.jp/Genomeseq.html.

FEATURES
 source
 1..166453
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="2"
 /clone="OSUNBA0064G16"
 338..3368
 /gene="OSUNBA0064G16.1"
 join(<338..475,2574..2681,2803..2943,3030..3203,
 3297..3368)
 /gene="OSUNBA0064G16.1"
 /note="start and end point are not identified"
 join(338..475,2574..2681,2803..2943,3030..3203,3297..3368)
 /gene="OSUNBA0064G16.1"
 /note="Predicted by GeneMark.hmm etc."
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="BADJ3231.1"
 /db_xref="GI:45736185"
 /translation="MSVFAIIIGYTAIDMTDIVITINISNIFITAYSTAIRDLR
 HHEHYKDLHPLPIVIAETLITDPTRELPLMLOMKAGSRMISWMSGTADPATL
 FLRYLYNREANLVEYESTFASRPVPLVLRKMSATWPTPTIRLMCLDEM
 ONAGSYDCCRLKLDLHSGIMSHLQGVVSDVLSLGGKMGQSN"
 join(4319..4454,6173..6241,6337..6636,6750..6851,
 7006..7200,7298..7585,7863..8049,9699..9745,9903..10264)
 /gene="OSUNBA0064G16.2"
 join(4319..4454,6173..6241,6337..6636,6750..6851,
 7006..7200,7298..7585,7863..8049,9699..9745,9903..10264)
 /gene="OSUNBA0064G16.2"
 /note="supported by full-length cDNA(s): AK063330"
 join(6182..6241,6337..6636,6750..6851,7006..7200,
 7298..7585,7863..8049,9699..9745,9903..9995)

gene
 mRNA
 CDS
 gene
 mRNA
 CDS

/gene="OSJNBa0064G16.2"
/note="contains full-length cDNA (s): AK063330"
/codon_start=1
/product="putative sucrose-phosphatase"
/protein_id="BAD1323.1"
/db_xref="GI:45736187"
/translation="MDKLNGKSRRLMIVSDLHTVWDHDEENLILRFGLMEAYVCO
DSLIVFSGRSPITLYMELRKREKPMHTPTITMSTGETTGEENVPDQGVYLNKKA
DRNIVESTANVSEIKLQVSEORPHKVSFTVMSKAOEVIKSLSEKLEKGLVTKIT
YSGQDDLVLPQAGKGOALVYLKQLSSCGKPPNNTLACDGSNDALFSPGVHGV
MVNSAOELLQWYSENAKDNPKIITHATERCAAGIIQAIHGFKLGPVNSPRVDEPVYK
ENPKPTDAVVKFVLYEKMRAVEPKSDSTVOYKNTTHANGVITHAGLECSLHAS
IDALGSCYGDGKKYKRAVDRIVASOCSSGMLVRFMLWELBGVWSCCLTSLALNA
KPTPEGRVYTHIKTKLKGYSADBOGSKL"
complement (8292..8879)
/gene="OSJNBa0064G16.3"
complement (8292..8879)
/gene="OSJNBa0064G16.3"
/note="probably inactive due to 5' exon missing in CDS
pseudogene, callose synthase"
pseudogene
/pseudo
13379..20883
/gene="OSJNBa0064G16.4"
join(13379..13733,14288..14362,14781..14896,15076..15190,
15281..15409,15730..15777,16290..16471,16549..16745,
16837..17123,17310..17459,17554..18015,18122..18271,
18533..18608,18772..18899,19165..19356,19480..19611,
19688..19810,20049..20267,20439..20883)
/gene="OSJNBa0064G16.4"
/note="supported by full-length cDNA (s): AK070600"
join(13569..13733,14288..14362,14781..14896,15076..15190,
15281..15409,15730..15777,16290..16471,16549..16745,
16837..17123,17310..17459,17554..18015,18122..18271,
18533..18608,18772..18899,19165..19356,19480..19611,
19688..19810,20049..20267,20439..20600)
/gene="OSJNBa0064G16.4"
/note="contains full-length cDNA (s): AK070600"
/codon_start=1
/product="cyclin-like protein"
/protein_id="BAD1323.1"
/db_xref="GI:45736188"
/translation="MGSVGLNDIVSGVSRKVLPAACGLCYFCPLRARSPVRYK
KIADIPATODEPERRIRIKCEYVARNRHPRPKITAYIEQRCYKELRBOQVPRK
VVVILIRKLVSCKQKPLASLSLIICTLDTRDMRIIGETLFDTVSQVDS
TYQNLBELVPKCELAQIVAEKDKMLRASTQALASAMWIGEBSEHISADNVY
QVLESYVLQMQNDINDSEAPGRWYEOVLKAGNATISRIIPWKSIVDGELHP
AEDADENFMGRVCVHNAKLSREATRFVLSLFRHFDNNSMSSONTLAFGLD
MOIMENOGNIDIMISILVKLHKKVLRQPEMQLAVEVYIASLAEGRASAAAT
GAISDLIRHMKTLHVALGSRDLEVIKMDLRAVADCEITQLSKYDAGPVLDMS
VMAISRTPLVATITTSVTRKQITISIPNLSIRNVKVFPEALPHQLLMVRDHS
TRVSAHRIFSVVLVSSVSPSKSTPNQLVHDIKTLASAVSFSSALFPKLR
DKSEPREKPDGSMNRSLSHADNDTSYKWDSSRSRSHKVPNFSKRVASLSLKS
PMSEKPCONTASBCESESTLLRLSSROATLSSLYIAQASIPNTKONVLAHTY
SILLFSGSKASIFEALPSFOVAFSLMSVLEGTDLIPRRSLFLTASIMFSS
RAHVAPLPIPKSMNERTMDPIHLVQDKLQAVYDCSEETGSEPDNNALKSLS
AVELTQOSRESASTNNNRDLPDSLQTRQSLDSFSDMCPSALFFELVTR
NPGQDESSNQEDVLINMANDTGEYVENTATTSVSPANLILGIDELLESVTLA
PSQTRCSVSSTAPNIPKEMTNOCEVLSMEKQCSVYLSLKHKQSGVSLINDADT
GAHAISDDQNTNPLQSLDGYPKYVADGALQVADVDVQOQFLKLPPASPYDTFL
KAAGC"
22576..22938
/gene="OSJNBa0064G16.5"
22576..22938
/gene="OSJNBa0064G16.5"
/note="start and end point are not identified"
22576..22938
/gene="OSJNBa0064G16.5"
/note="predicted by RGENESH etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD1323.1"
/db_xref="GI:45736189"

/translation="MGVPRAPRRKAAAGSESRALLVGGGSEEAAPVGPVAVY
GAERKRFVPTSYLRQPAFRGLMELADDFGFAOEGGILRUPCREEDFOATVALDARR
RPAAGGAIWSTVVARSL"
complement (24680..25238)
/gene="OSJNBa0064G16.6"
complement (join(24680..24761,24856..25238))
/gene="OSJNBa0064G16.6"
/note="start and end point are not identified"
complement (join(24680..24761,24856..25238))
/gene="OSJNBa0064G16.6"
/codon_start=1
/product="auxin-induced protein-like"
/protein_id="BAD1323.1"
/db_xref="GI:45736190"
/translation="MGSQGRASSNKIRIDIVRLHQLIKRMKRAALAPKAGNNNGGA
SVPKGFPAVCVEEMERFVITPYLGMAFEQLRKAEERGFQHEALRIPCVVEVF
EGILRIYGRDKERKAMCYSSSEHILTEEEFVGLVLCYFQEDVNMNSHF"
complement (28164..31674)
/gene="OSJNBa0064G16.7"
complement (join(28164..29428,31041..31674))
/gene="OSJNBa0064G16.7"
/note="start and end point are not identified"
complement (join(28164..29428,31041..31674))
/gene="OSJNBa0064G16.7"
/note="similar to Oryza sativa chromosome 8,
OJ1449.C01.27"
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD1323.1"
/db_xref="GI:45736191"
/translation="MRRSDKIVGATLYDPASHVAVLPSMWYKFMWOSFAVGDLL
YVNETGRALGCHPQRLRIYVNGERVRKAGMVALPFRGRATVYEHGILMPLSLA
DDEGLWMDLSASTVADQPPPAQPCGVPVEPAPGTGTVHVGGLCVAKLY
MVAPRGSCPCCGEEDMKFMFLITVEVARGGRGELIVRKSCRYSGANGVNV

Query Match 84.8%; Score 17.8; DB 8; Length 166453;
Best Local Similarity 90.5%; Pred. No. 3e+02; 2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCCGATGATGATTACCTCC 21
Db 114005 TCCGATGATGATTACCAACC 114025

RESULT 10
AC079447 175567 bp DNA linear PRI 09-JAN-2002
LOCUS
DEFINITION Homo sapiens BAC clone Rpl1-111H13 from 2, complete sequence.
AC079447
AC079447.4 GI:15145541
VERSION
KEYWORDS
HUG.
SOURCE
ORGANISM Homo sapiens (human)
Homosapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 175567)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
2 (bases 1 to 175567)
Waterston,R.H.
Submitted (01-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Farway, St. Louis,
MO 63108, USA
4 (bases 1 to 175567)
Waterston,R.H.

TITLE Direct Submission
JOURNAL Submitted (09-AUG-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 175567)
AUTHORS Waterston,R.
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 9, 2001 this sequence version replaced g114550319.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics

Center project name: H_NH011H13

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osogawa,K., Moon,P.Y., Zhao,B., Frengen,B.,
Tateno,M., Calanese,J.V. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBAC3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-38C17, the clone sequenced
to the right is RP11-527J8, 2000 bp overlap. Actual start of this
clone is at base position 1 of RP11-11H13; actual end is at base
position 5330 of RP11-527J8.

Data from AC023965 and AC018690 was used to finish this clone,
AC079447.

FEATURES
SOURCE
1. 175567 location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-11H13"
/clone_lib="RPCI-11"
misc_feature
1. .795 /note="match to EST AL581351 (NID:g12948262) "
misc_feature
1. .552 /note="match to EST BG028419 (NID:g12417513) "
misc_feature
1. .470 /note="similar to Homo sapiens EST BG687463
(NID:g13918860) "
misc_feature
1. .463

misc_feature
1. .312 /note="match to EST A1148460 (NID:g3676929) qc39e04.x1"
repeat_region
3. .123 /note="match to EST AM391243 (NID:g6896006) "
misc_feature
15. .763 /rpt_family="L2"
misc_feature
21. .457 /note="match to EST BE620779 (NID:g9891717) "
misc_feature
42. .792 /note="match to EST AV703935 (NID:g10733284) "
misc_feature
/note="match to EST BG742002 (NID:g14052655) "
misc_feature
211. .970 /note="match to EST BG284097 (NID:g13034702) "
misc_feature
214. .393 /note="match to EST W02815 (NID:g1274948) za07d03.r1"
misc_feature
250. .817 /note="match to EST A1309762 (NID:g4004633) qo75a08.x1"
misc_feature
254. .732 /note="match to EST AA459888 (NID:g2184795) zx65h10.r1"
misc_feature
353. .711 /note="match to EST AA459888 (NID:g2184795) zx65h10.r1"
misc_feature
358. .930 /note="similar to Homo sapiens EST R28040 (NID:g784175)
yh6b03.r1"
misc_feature
/note="match to EST W37953 (NID:g1319547) zc13e02.r1"
misc_feature
419. .798 /note="similar to Homo sapiens EST R28262 (NID:g784397)
yh6b03.r1"
misc_feature
469. .845 /note="match to EST T71834 (NID:g686355) yc62g12.r1"
misc_feature
515. .785 /note="match to EST AV703935 (NID:g10733284) "
misc_feature
526. .998 /note="match to EST AA633739 (NID:g2556953) ag87f03.r1"
misc_feature
527. .796 /note="match to EST BG214283 (NID:g13740304) "
misc_feature
529. .1134 /note="match to EST BE256300 (NID:g9126854) "
misc_feature
863 /note="match to EST T71834 (NID:g686355) yc62g12.r1"
misc_feature
883. .884 /note="similar to Homo sapiens EST R28262 (NID:g784397)
yh6b03.r1"
misc_feature
989. .1680 /note="match to EST BE535937 (NID:g9764582) "
repeat_region
1019. .1316 /rpt_family="Alu"
misc_feature
1220. .1944 /note="match to EST BG495526 (NID:g12457041) "
repeat_region
1523. .2003 /rpt_family="MAIR"
misc_feature
1617. .2399 /note="match to EST BG613322 (NID:g1364693) "
misc_feature
1729. .2117 /note="match to EST AA773220 (NID:g2824791) af78e04.r1"
misc_feature
1877. .2332 /note="match to EST AW297722 (NID:g6704358) "
misc_feature
2238. .2239 /note="match to EST A1830655 (NID:g5451326) wj10d06.x1"
misc_feature
2455. .2931 /note="match to EST A1830655 (NID:g5451326) wj10d06.x1"
misc_feature
2456. .2634 /note="match to EST AW798659 (NID:g7850542) "
misc_feature
2847. .3164 /note="match to EST AA971709 (NID:g3146999) op95b11.r1"
repeat_region
3043. .3184 /rpt_family="Alu"
misc_feature
3099. .3330 /note="match to EST BF811752 (NID:g12140940) "
misc_feature
3525. .4128 /note="match to EST AL135413 (NID:g6603600) "
misc_feature
3632. .3633 /note="match to EST BF811752 (NID:g12140940) "
repeat_region
3747. .4023

```
repeat_region      /rpt_family="Alu"
                    4330..4485
repeat_region      /rpt_family="ERV1"
                    4528..4827
repeat_region      /rpt_family="ERV1"
                    4994..5511
repeat_region      /rpt_family="ERV1"
                    5891..5918
repeat_region      /rpt_family="AT-rich"
                    5905..6091
repeat_region      /rpt_family="L1"
                    6092..6352
repeat_region      /rpt_family="Alu"
                    6339..6372
repeat_region      /rpt_family="AT-rich"
                    6353..7447
repeat_region      /rpt_family="L1"
                    7439..7510
repeat_region      /rpt_family="CT-rich"
                    7489..7724
repeat_region      /rpt_family="Alu"
                    7727..8263
repeat_region      /rpt_family="L1"
                    8260..8295
repeat_region      /rpt_family="(T)n"
                    8266..8523
repeat_region      /rpt_family="Alu"
                    8266..8523
```

```
Query Match      84.8%; Score 17.8; DB 9; Length 175567;
Best Local Similarity 90.5%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1 TCCGAGATGATTAATCTCC 21
Db      86970 TCCGCGTGAATGATTAATCTCC 86990
```

```
RESULT 11
AC092002/c      195744 bp      DNA      linear      HTG 07-SEP-2001
LOCUS      Homo sapiens chromosome RP11-364D20, WORKING DRAFT
DEFINITION      AC092002
ACCESSION      AC092002
VERSION      HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 195744)
AUTHORS      Waterston,R.H.
TITLE      The sequence of Homo sapiens clone
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 195744)
AUTHORS      Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (11-JUN-2001) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT      On Sep 7, 2001 this sequence version replaced gi:14336624.
```

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0364D20
----- Summary Statistics -----
Sequencing vector: M13; 12%
Sequencing vector: plasmid; 88%
Chemistry: Dye-primer RT; 6% of reads
Chemistry: Dye-terminator Big Dye; 94% of reads
```

```
Assembly program: Phrap: version 0.990119
Consensus quality: 189590 bases at least Q40
Consensus quality: 191694 bases at least Q30
Consensus quality: 192828 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
* 1 11257: contig of 11257 bp in length
* 11258: gap of unknown length
* 11358: contig of 16616 bp in length
* 27973: gap of unknown length
* 28073: gap of unknown length
* 28074: contig of 21597 bp in length
* 49671: gap of unknown length
* 49770: gap of unknown length
* 49771: contig of 14598 bp in length
* 64369: gap of unknown length
* 64469: gap of unknown length
* 83993: contig of 19525 bp in length
* 84093: gap of unknown length
* 84094: contig of 24530 bp in length
* 108623: gap of unknown length
* 108624: gap of unknown length
* 137943: contig of 29220 bp in length
* 137944: gap of unknown length
* 138044: gap of unknown length
* 184871: contig of 46828 bp in length
* 184972: gap of unknown length
* 184973: gap of unknown length
* 187036: contig of 2065 bp in length
* 187037: gap of unknown length
* 187137: gap of unknown length
* 187137 195744: contig of 8608 bp in length.
```

FEATURES

source

```
1..195744
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="RP11-364D20"
1..11257
/note="assembly_name:Contig10"
11358..27973
/note="assembly_name:Contig11"
clone_end:T7
vector_side:left"
28074..49670
/note="assembly_name:Contig12"
49771..64368
/note="assembly_name:Contig13"
64469..83993
/note="assembly_name:Contig14"
clone_end:SP6
vector_side:left"
84094..108623
/note="assembly_name:Contig15"
108724..137943
/note="assembly_name:Contig16"
138044..184871
/note="assembly_name:Contig17"
184972..187036
/note="assembly_name:Contig18"
187137..195744
/note="assembly_name:Contig19"
```

ORIGIN

```
Query Match      84.8%; Score 17.8; DB 2; Length 195744;
Best Local Similarity 90.5%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1 TCCGAGATGATTAATCTCC 21
Db      16856 TCCGCGTGAATGATTAATCTCC 16836
```

RESULT 12
AC112285/c
LOCUS
DEFINITION
AC112285
AC112285
AC112285.4 GI:30580789
HTG: HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eutheria; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 247746)
Munry, D., Marle, Metzker, M., Lee, A., Branson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Blawie, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Deigado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Drapper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gargese, J., Geer, K., Gill, R., Grady, M., Guerra, W., Guerra, W.,
Guarinate, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howell, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kovacs, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, L., Louised, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindrasekaran, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Matsushita, M., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, N., Munday, N., Murphy, M., Nair, L.,
Narkewicz, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwokwelen, O., Okunolu, G., Olampunase, A., Pal, S., Parks, K.,
Paetel, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poldexter, A., Popovic, D., Primus, B., Pu, L., L.,
Puzo, B., Qutroz, J., Rachlin, R., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Rodkey, T., Rojce, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, M., Saverly, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Sivarsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
Sned, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Tjio, J., Umani, K.,
Vallas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wiczyska, R., Woodson, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 247746)
Worley, K. C.
Direct Submission
Submitted (21-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 247746)

AUTHORS
TITLE
JOURNAL

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:23664473.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: G1XQ
Center clone name: CH230-97H21
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 236310 bases at least Q40
Consensus quality: 236673 bases at least Q30
Consensus quality: 240367 bases at least Q20
Estimated insert size: 248967; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbankdraft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 246339: contig of 246339 bp in length
* 246340 247746: gap of unknown length
* 246440 247746: contig of 1307 bp in length.
Location/Qualifiers
1. 247746
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-97H21"
1. 11764
/note="wgs_end-extension"
/clone_end="5p6"
6257. 6913
/note="clone boundary"
clone_end="5p6"
site:ECORI
end_sequence="BH282614"
150844. 151647
/note="clone boundary"
clone_end="T7"
site:ECORI
end_sequence="BH282612"

FEATURES

source
1. 247746
Location/Qualifiers
1. 247746
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-97H21"
1. 11764
/note="wgs_end-extension"
/clone_end="5p6"
6257. 6913
/note="clone boundary"
clone_end="5p6"
site:ECORI
end_sequence="BH282614"
150844. 151647
/note="clone boundary"
clone_end="T7"
site:ECORI
end_sequence="BH282612"

misc_feature

misc_feature
/note="wgs_end-extension"
/clone_end="5p6"
6257. 6913
/note="clone boundary"
clone_end="5p6"
site:ECORI
end_sequence="BH282614"
150844. 151647
/note="clone boundary"
clone_end="T7"
site:ECORI
end_sequence="BH282612"

misc_feature

misc_feature
/note="wgs_end-extension"
/clone_end="5p6"
6257. 6913
/note="clone boundary"
clone_end="5p6"
site:ECORI
end_sequence="BH282614"
150844. 151647
/note="clone boundary"
clone_end="T7"
site:ECORI
end_sequence="BH282612"

misc_feature

misc_feature
/note="wgs_end-extension"
/clone_end="5p6"
6257. 6913
/note="clone boundary"
clone_end="5p6"
site:ECORI
end_sequence="BH282614"
150844. 151647
/note="clone boundary"
clone_end="T7"
site:ECORI
end_sequence="BH282612"

ORIGIN

Query Match 84.8%; Score 17.8; DB 2; Length 247746;
Best Local Similarity 90.5%; Pred. No. 3.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 TCCGATGATGATTTACTCTC 21

DB 154435 TCAGATGATGCTTTACTCC 154415

|||||

RESULT 13
AC128496/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-444D9, WORKING DRAFT SEQUENCE, 3
unordered pieces.

AC128496 291544 bp DNA linear HTG 19-NOV-2002
AC128496.3 GI:25085139
VERSION HTG, HTGS_PHASE1, HTGS_DRAFT, HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 291544)
Murny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooke,S., Amin,A., Angiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,D.K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,U., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garra,M.,
Gerogeorgis,B., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunnarsson,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulik,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kovis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Louisedge,H., Lozard,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmood,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinley,S., McLeod,M.P., McNeill,T.Z., Meenen,B.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwokolemeh,O., Okunom,G., Olarunpugoon,A., Pal,S., Parke,K.,
Pateranak,S., Paul,H., Perez,A., Popovic,D., Primus,B., Pu,L.,
Plopper,F., Poindexter,A., Popovic,D., Primus,B., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Rejzler,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,R., Richards,S., Riggs,F.,
Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shateman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,C., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Uemari,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,O., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wiczek,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,J., Zhao,S., Zhao,S., Yoon,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 291544)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23907748.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYSW
Center clone name: CH230-444D9
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 187201 bases at least Q40
Consensus quality: 189849 bases at least Q30
Consensus quality: 191486 bases at least Q20
Estimated insert size: 193264; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 288506: contig of 288506 bp in length
* 288507 288606: gap of unknown length
* 288607 289949: contig of 1343 bp in length
* 289950 290049: gap of unknown length
* 290050 291544: contig of 1495 bp in length.

Location/Qualifiers
1. 291544
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-444D9"
1. 1841
/note="wgs end-extension
clone end: 77"
6777..7677
/note="clone boundary
clone end: 77"
site:
end_sequence: BZ198832"
complement(187560..188645)
/note="clone boundary
clone end: 5p6
site:
end_sequence: BZ198833"
189481..192835
/note="wgs_end-extension

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

misc_feature clone_end:Sp6"
195005..196648
/note="wgs_end_extension
clone_end:Sp6"

ORIGIN

Query Match 84.8%; Score 17.8; DB 2; Length 291544;
Best Local Similarity 90.5%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCCGATGATGATTACTCC 21
|||||
Db 10095 TCAGATGATGTTTACTCC 10075
|||||

RESULT 14
AC149573 320731 bp DNA linear HTG 06-OCT-2004
LOCUS Medicago truncatula clone mth2-119g4, WORKING DRAFT SEQUENCE, 52
DEFINITION unordered pieces.
AC149573
AC149573.8 GI:53828780
VERSION HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS Medicago truncatula (barrel medic)
SOURCE Medicago truncatula
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales;
Medicago.
1 (bases 1 to 320731)
Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
and Roe,B.A.
Medicago truncatula BAC Clone mth2-119g4
2 (bases 1 to 320731)
Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
and Roe,B.A.
Direct Submission
Submitted (08-JUN-2004) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Oct 6, 2004 this sequence version replaced gi:53793764.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

* NOTE: This is a 'working draft' sequence. It currently
* consists of 52 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2120: contig of 2120 bp in length
* 1 2121 2220: gap of unknown length
* 2221 4377: contig of 2157 bp in length
* 4378 4477: gap of unknown length
* 4478 6499: contig of 2022 bp in length
* 6500 8659: gap of unknown length
* 8660 8658: contig of 2059 bp in length
* 8659 8758: gap of unknown length
* 8759 11276: contig of 2518 bp in length
* 11277 11376: gap of unknown length
* 11377 13990: contig of 2614 bp in length

13991 14090: gap of unknown length
* 14091 16332: contig of 2242 bp in length
* 16333 16432: gap of unknown length
* 16433 18500: contig of 2068 bp in length
* 18501 18600: gap of unknown length
* 18601 21151: contig of 2551 bp in length
* 21152 21251: gap of unknown length
* 21252 23489: contig of 2238 bp in length
* 23490 23589: gap of unknown length
* 23590 26707: contig of 3118 bp in length
* 26708 26807: gap of unknown length
* 26808 30025: contig of 3218 bp in length
* 30026 30125: gap of unknown length
* 30126 33133: contig of 3008 bp in length
* 33134 33233: gap of unknown length
* 33234 35396: contig of 2163 bp in length
* 35397 35496: gap of unknown length
* 35497 38285: contig of 2789 bp in length
* 38286 42217: gap of unknown length
* 42218 42317: gap of unknown length
* 42318 44932: contig of 2615 bp in length
* 44933 45032: gap of unknown length
* 45033 48055: contig of 3023 bp in length
* 48056 48155: gap of unknown length
* 48156 50571: contig of 2416 bp in length
* 50572 50671: gap of unknown length
* 50672 52786: contig of 2115 bp in length
* 52787 52886: gap of unknown length
* 52887 55014: contig of 2128 bp in length
* 55015 55114: gap of unknown length
* 55115 58456: contig of 3342 bp in length
* 58457 58556: gap of unknown length
* 58557 60574: contig of 2018 bp in length
* 60575 60674: gap of unknown length
* 60675 64933: contig of 4259 bp in length
* 64934 65033: gap of unknown length
* 65034 68622: contig of 3589 bp in length
* 68623 73490: gap of unknown length
* 73491 75590: contig of 4768 bp in length
* 75591 77278: gap of unknown length
* 77279 80719: contig of 3688 bp in length
* 80720 80819: gap of unknown length
* 80820 85277: contig of 4458 bp in length
* 85278 85377: gap of unknown length
* 85378 92306: contig of 6929 bp in length
* 92307 92406: gap of unknown length
* 92407 97236: contig of 4830 bp in length
* 97237 102561: gap of unknown length
* 102562 102661: contig of 5225 bp in length
* 102662 106887: gap of unknown length
* 106888 106987: contig of 4226 bp in length
* 106989 112013: gap of unknown length
* 112014 112113: contig of 5026 bp in length
* 112114 115962: gap of unknown length
* 115963 117062: contig of 4849 bp in length
* 117063 122443: gap of unknown length
* 122444 123443: contig of 5181 bp in length
* 123445 131655: gap of unknown length
* 131656 131755: contig of 9312 bp in length
* 131756 137837: gap of unknown length
* 137838 137937: contig of 6082 bp in length
* 137938 144111: gap of unknown length
* 144112 144211: contig of 6174 bp in length
* 144212 153511: gap of unknown length
* 153512 153611: contig of 9300 bp in length
* 153612 159483: gap of unknown length
* 159484 159583: contig of 5872 bp in length
* 159584 168698: gap of unknown length
* 168699 168798: contig of 9115 bp in length


```
* 168799 178460: contig of 9662 bp in length
* 178461 178560: gap of unknown length
* 178561 189146: contig of 10586 bp in length
* 189147 189246: gap of unknown length
* 189247 205996: contig of 16750 bp in length
* 205997 206097: gap of unknown length
* 206097 221512: contig of 15416 bp in length
* 221513 235904: gap of unknown length
* 235905 236005: gap of unknown length
* 236005 248143: contig of 12138 bp in length
* 248143 248242: gap of unknown length
* 248243 263641: contig of 15399 bp in length
* 263642 263741: gap of unknown length
* 263742 280695: contig of 16954 bp in length
* 280696 280795: gap of unknown length
* 280796 301898: contig of 21103 bp in length
* 301899 301998: gap of unknown length
* 301999 320731: contig of 18733 bp in length
Location/Qualifiers
1..320731
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-119g4"
/clone_lib="Medicago truncatula BAC library H2"
```

ORIGIN

```
Query Match      84.8%; Score 17.8; DB 2; Length 320731;
Best Local Similarity 90.5%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY      1 TCCGATGATGATTACCTC 21
Db      95671 TCCGATGATGATTACCTC 95691
```

```
RESULT 15
ATTU90428/c      1286 bp DNA linear PLN 25-MAR-1997
DEFINITION      Arabidopsis thaliana blue copper-binding protein II gene, complete cds.
```

```
ACCESSION      U90428
VERSION        U90428.1 GI:1905999
```

```
KEYWORDS
SOURCE
ORGANISM      Arabidopsis thaliana (thale cress)
```

```
REFERENCE
AUTHORS      Arabidopsis thaliana
TITLE        Arabidopsis thaliana
JOURNAL      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
```

```
1 (bases 1 to 1286)
Kim,C.H., Cho,Y.H. and Hong,Y.-N.
Direct Submission
Submitted (22-FEB-1997) Department of Biology, College of Natural Science, Seoul National University, 56-1 Shillim-Dong, Kwanak-Ku, Seoul 151-742, Korea
```

```
FEATURES
SOURCE
Location/Qualifiers
```

```
1..1286
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
join(160..331,774..1207)
note="BCB II"
/codon_start=1
/product="blue copper-binding protein II"
/protein_id="AAB50232.1"
/db_xref="GI:1906000"
```

CDS

```
/translation="MAAMGSKMAVAAATALLVLTIVPGAVAVTYTTEMTGVDSG
WATGTRPVGILIEFKYSSHTVDVVDKAGDGDASSSTENHSDGDTKIDLVGIN
YFICSTPGHCSLNGMKLAIVNVAGSADLRTPSPSPGTPPTPESSPGSGSPPTT
PTPGASTSPPPKASGASGXVMSYLVGVSMVLGYGLMM"
```

ORIGIN

```
Query Match      82.9%; Score 17.4; DB 8; Length 1286;
Best Local Similarity 94.7%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 TCCGATGATGATTACCT 19
Db      348 TCCGATGATGATTACCT 330
```

```
Search completed: November 7, 2005, 06:41:41
Job time : 837.255 secs
```

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:04:21 ; Search time 242.17 Seconds
(without alignments)
513.336 Million cell updates/sec

Title: US-10-777-131a-1

Perfect score: 21

Sequence: 1 tccgagcagcattactctcc 21

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	8	ABZ79927 Mycobacte
2	21	100.0	823	8	ABZ79951 Mycobacte
3	21	100.0	3002	13	AD575303 Mycobacte
4	21	100.0	110000	4	AA199682_12 Confination (13 o
5	21	100.0	110000	4	AA199683_12 Confination (13 o
6	17.4	82.9	1275	8	ACA47360 Prokaryot
7	17	81.0	397	3	AAA15451 Aaa15451 DNA seque
8	17	81.0	110000	10	AD577343_03 Continuation (4 of
9	16.8	80.0	131239	12	AD021179 Human sof
10	16.2	77.1	225	9	AC192883 Human mic
11	16.2	77.1	288	9	AC11729 Arabidops
12	16.2	77.1	304	9	ADA58045 Maize trl
13	16.2	77.1	304	12	ADP60670 Maize car
14	16.2	77.1	316	9	ADA57946 Maize trl
15	16.2	77.1	316	12	ADP60571 Maize car
16	16.2	77.1	325	12	ADP60572 Maize car
17	16.2	77.1	325	12	ADP60572 Maize car
18	16.2	77.1	325	12	ADP60572 Maize car
19	16.2	77.1	325	12	ADP60572 Maize car
20	16.2	77.1	466	3	AA134481 Arabidops

21	16.2	77.1	1116	8	ACA20952 Prokaryot
22	16.2	77.1	1131	3	ADA30615 DNA encod
23	16.2	77.1	1221	9	AA134666 Zea maye
24	16.2	77.1	1329	8	ACA30420 Prokaryot
25	16.2	77.1	3435	6	AB266862 Arabidops
26	16.2	77.1	4647	6	ABA90426 Drosophi1
27	16.2	77.1	5316	4	AB111843 Drosophi1
28	16.2	77.1	7499	4	AB111842 Drosophi1
29	16.2	77.1	14775	4	AB117452 Drosophi1
30	16.2	77.1	110000	12	ADN46845_06 Continuation (7 of
31	16.2	77.1	110000	12	ADN47591_14 Continuation (7 of
32	16.2	77.1	110000	12	ADN46123_06 Continuation (7 of
33	16.2	77.1	110000	12	ADN47209_14 Continuation (7 of
34	16.2	77.1	110000	12	ADN46464_06 Continuation (7 of
35	16.2	77.1	110000	12	ADN47960_14 Continuation (15 o
36	16	76.2	1080	8	ACA31076 Prokaryot
37	16	76.2	3510	3	AA165299 Maize sca
38	15.8	75.2	434	12	ADQ20901 Human sof
39	15.8	75.2	557	12	ADQ24851 Human sof
40	15.8	75.2	1215	13	ADT41714 Bacteri
41	15.8	75.2	4398	6	ADH32179 Yeast sma
42	15.8	75.2	110000	3	AA122305 Arabidops
43	15.8	75.2	110000	3	AA122305 Arabidops
44	15.8	75.2	116624	2	AAV52850 Human eya
45	15.4	73.3	216	2	AA112103 Human bia

ALIGNMENTS

RESULT 1	ABZ79927	standard; DNA; 21 BP.
ID	ABZ79927	standard; DNA; 21 BP.
XX	ABZ79927	
AC	ABZ79927	
XX	ABZ79927	
DT	19-MAY-2003	(first entry)
XX	19-MAY-2003	
DE	Mycobacterium tuberculosis mutT2-1 PCR primer SEQ ID NO.1.	
XX	Mycobacterium tuberculosis; mutT2; alkA; ogt; RV3908; mutY; RV3909;	
KW	detection; multidrug resistance; multiple drug resistance; MDR;	
XX	infection; PCR primer; ss.	
OS	Mycobacterium tuberculosis.	
XX	Synthetic.	
PN	WO2003016562-A2.	
XX	27-FEB-2003.	
PD	27-FEB-2003.	
XX	14-AUG-2002; 2002WO-EP009679.	
PF	14-AUG-2002; 2002WO-EP009679.	
XX	14-AUG-2001; 2001US-0311824P.	
PR	21-AUG-2001; 2001US-0313523P.	
XX	(INSP) INST PASTEUR.	
PA	(INSP) INST PASTEUR.	
XX	Gicquel B;	
PI	Gicquel B;	
XX	WPI, 2003-256711/25.	
DR	WPI, 2003-256711/25.	
XX		
PT	Predicting the epidemic character of a Mycobacterium tuberculosis isolate	
XX	and/or the acquisition of multiple drug resistance (MDR) by the isolate	
PT	by detecting an alteration in the DNA repair system of the isolate.	
XX		
PS	Claim 32; Page 16; 83pp; English.	
XX		
CC	The present invention describes a method for predicting the epidemic	
CC	character of a Mycobacterium tuberculosis isolate and/or a selective	
CC	advantage to be maintained in the host and/or the acquisition of multiple	
CC	drug resistance (MDR) by the isolate comprising detecting an alteration	
CC	in the DNA repair system of the isolate. Also described: (1) detecting a	

```
CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a
CC polymnucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)
CC an Escherichia coli strain containing the plasmid pMYC2501; and (5)
CC detecting in a patient infected by Mycobacterium tuberculosis a higher
CC risk of being unable to eliminate the bacillus or of developing MDR
CC tuberculosis. The method is useful for predicting the epidemic character
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to
CC be maintained in the host and/or the acquisition of MDR by the isolate.
CC The present sequence represents a PCR primer for M. tuberculosis mutT2,
CC which is used in the exemplification of the present invention
XX
SQ Sequence 21 BP; 4 A; 6 C; 4 G; 7 T; 0 U; 0 Other;
Query Match 100.0%; Score 21; DB 8; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TCCGGATGATGATTACTCTCC 21
DB 1 TCCGGATGATGATTACTCTCC 21
RESULT 2
ABZ79951
ID ABZ79951 standard; DNA; 823 BP.
XX
AC ABZ79951;
XX
DT 19-MAY-2003 (first entry)
XX
DE Mycobacterium tuberculosis mutT2 nucleotide sequence SEQ ID NO.19.
XX
KW Mycobacterium tuberculosis; mutT2; alkA; ogt; RV3908; mutY; RV3909;
KW detection; multidrug resistance; multiple drug resistance; MDR;
KW infection; gene; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO2003016562-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002MO-EP009679.
XX
PR 14-AUG-2001; 2001US-0311824P.
XX
PR 21-AUG-2001; 2001US-0313523P.
XX
PA (INSP ) INST PASTEUR.
XX
PI Glacquel B;
XX
DR WPI; 2003-256711/25.
XX
PT Predicting the epidemic character of a Mycobacterium tuberculosis isolate
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate
PT by detecting an alteration in the DNA repair system of the isolate.
XX
PS Disclosure; Fig 4A; 83pp; English.
XX
CC The present invention describes a method for predicting the epidemic
CC character of a Mycobacterium tuberculosis isolate and/or a selective
CC advantage to be maintained in the host and/or the acquisition of multiple
CC drug resistance (MDR) by the isolate comprising detecting an alteration
CC in the DNA repair system of the isolate. Also described: (1) detecting a
CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a
CC polymnucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)
CC an Escherichia coli strain containing the plasmid pMYC2501; and (5)
CC detecting in a patient infected by Mycobacterium tuberculosis a higher
CC risk of being unable to eliminate the bacillus or of developing MDR
CC tuberculosis. The method is useful for predicting the epidemic character
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to
CC be maintained in the host and/or the acquisition of MDR by the isolate.
CC The present sequence represents a M. tuberculosis mutT2 nucleotide
```

```
CC sequence, which is used in the exemplification of the present invention
XX
SQ Sequence 823 BP; 144 A; 253 C; 277 G; 149 T; 0 U; 0 Other;
Query Match 100.0%; Score 21; DB 8; Length 823;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TCCGGATGATGATTACTCTCC 21
DB 5 TCCGGATGATGATTACTCTCC 25
RESULT 3
ADS75303
ID ADS75303 standard; DNA; 3002 BP.
XX
AC ADS75303;
XX
DT 16-DEC-2004 (first entry)
XX
DE Mycobacterium tuberculosis nargHJI nitrate reductase operon Segid 1.
XX
KW nargHJI nitrate reductase operon; single nucleotide polymorphism; SNP;
KW detection method; ds.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT variation replace(1500..C)
FT /*tag= a
FT /*standard_name= "Single nucleotide polymorphism"
FT misc_feature 1717..1719
FT /*tag= b
FT /*note= "GTC start codon"
XX
PN WO2004083459-A1.
XX
PD 30-SEP-2004.
XX
PF 19-MAR-2004; 2004MO-EP002911.
XX
PR 20-MAR-2003; 2003DE-01013791.
XX
PA (ARTU-) ARTUS GBS MOLEKULARBIOLOGISCHE DIAGNOSTI.
XX
PI Bange F;
XX
DR WPI; 2004-728487/71.
XX
PT Specific detection of Mycobacterium tuberculosis, useful for differential
PT diagnosis of infection, by detecting specific polymorphism in the
PT promoter of the nitrate reductase operon.
XX
PS Claim 1; SEQ ID NO 1; 46pp; German.
XX
CC This invention relates to a novel method for the detection of
CC Mycobacterium tuberculosis (M. tuberculosis) occurring in a biological
CC sample. Specifically, it refers to the amplification of the nargHJI
CC nitrate reductase operon, including position -215 in the 5' to 3' reading
CC direction that occurs upstream of the GTC start codon of the operon. The
CC present invention describes detecting an M. tuberculosis specific
CC polymorphism at position -215. The polymorphism (C-215T) is detected by
CC hybridisation to specific probes such that M. tuberculosis can be
CC differentiated from M. bovis, M. bovis BCG, M. africanum and M. microti
CC by analysis of the melting curve. The method is used to detect M.
CC tuberculosis infection, especially to differentiate it from other species
CC of the M. tuberculosis complex, and furthermore provides rapid and
CC specific detection, and differentiation of M. tuberculosis. This
CC polymnucleotide sequence is the M. tuberculosis nargHJI nitrate reductase
CC operon containing the -215 SNP at position 1500 in this sequence of the
CC invention.
XX
```

SQ Sequence 3002 BP; 483 A; 923 C; 1013 G; 583 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 13; Length 3002;

Best Local Similarity 100.0%; Pred. No. 1,7; Mismatches 0; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCGATGATGATTTACTCTCC 21

DB 788 TCCGATGATGATTTACTCTCC 808

RESULT 4

AA199682_12
Continuation (13 of 45) of AA199682 from base 1200001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AA199682 Accession AA199682

WP	Fragment Name	Begin	End
WP	AA199682_00	1	110000
WP	AA199682_01	100001	210000
WP	AA199682_02	200001	310000
WP	AA199682_03	300001	410000
WP	AA199682_04	400001	510000
WP	AA199682_05	500001	610000
WP	AA199682_06	600001	710000
WP	AA199682_07	700001	810000
WP	AA199682_08	800001	910000
WP	AA199682_09	900001	1010000
WP	AA199682_10	1000001	1110000
WP	AA199682_11	1100001	1210000
WP	AA199682_12	1200001	1310000
WP	AA199682_13	1300001	1410000
WP	AA199682_14	1400001	1510000
WP	AA199682_15	1500001	1610000
WP	AA199682_16	1600001	1710000
WP	AA199682_17	1700001	1810000
WP	AA199682_18	1800001	1910000
WP	AA199682_19	1900001	2010000
WP	AA199682_20	2000001	2110000
WP	AA199682_21	2100001	2210000
WP	AA199682_22	2200001	2310000
WP	AA199682_23	2300001	2410000
WP	AA199682_24	2400001	2510000
WP	AA199682_25	2500001	2610000
WP	AA199682_26	2600001	2710000
WP	AA199682_27	2700001	2810000
WP	AA199682_28	2800001	2910000
WP	AA199682_29	2900001	3010000
WP	AA199682_30	3000001	3110000
WP	AA199682_31	3100001	3210000
WP	AA199682_32	3200001	3310000
WP	AA199682_33	3300001	3410000
WP	AA199682_34	3400001	3510000
WP	AA199682_35	3500001	3610000
WP	AA199682_36	3600001	3710000
WP	AA199682_37	3700001	3810000
WP	AA199682_38	3800001	3910000
WP	AA199682_39	3900001	4010000
WP	AA199682_40	4000001	4110000
WP	AA199682_41	4100001	4210000
WP	AA199682_42	4200001	4310000
WP	AA199682_43	4300001	4410000
WP	AA199682_44	4400001	4411529

Query Match 100.0%; Score 21; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 3; Mismatches 0; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCGATGATGATTTACTCTCC 21

DB 86397 TCCGATGATGATTTACTCTCC 86417

RESULT 5
AA199683_12

Continuation (13 of 44) of AA199683 from base 1200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AA199683 Accession AA199683

WP	Fragment Name	Begin	End
WP	AA199683_00	1	110000
WP	AA199683_01	100001	210000
WP	AA199683_02	200001	310000
WP	AA199683_03	300001	410000
WP	AA199683_04	400001	510000
WP	AA199683_05	500001	610000
WP	AA199683_06	600001	710000
WP	AA199683_07	700001	810000
WP	AA199683_08	800001	910000
WP	AA199683_09	900001	1010000
WP	AA199683_10	1000001	1110000
WP	AA199683_11	1100001	1210000
WP	AA199683_12	1200001	1310000
WP	AA199683_13	1300001	1410000
WP	AA199683_14	1400001	1510000
WP	AA199683_15	1500001	1610000
WP	AA199683_16	1600001	1710000
WP	AA199683_17	1700001	1810000
WP	AA199683_18	1800001	1910000
WP	AA199683_19	1900001	2010000
WP	AA199683_20	2000001	2110000
WP	AA199683_21	2100001	2210000
WP	AA199683_22	2200001	2310000
WP	AA199683_23	2300001	2410000
WP	AA199683_24	2400001	2510000
WP	AA199683_25	2500001	2610000
WP	AA199683_26	2600001	2710000
WP	AA199683_27	2700001	2810000
WP	AA199683_28	2800001	2910000
WP	AA199683_29	2900001	3010000
WP	AA199683_30	3000001	3110000
WP	AA199683_31	3100001	3210000
WP	AA199683_32	3200001	3310000
WP	AA199683_33	3300001	3410000
WP	AA199683_34	3400001	3510000
WP	AA199683_35	3500001	3610000
WP	AA199683_36	3600001	3710000
WP	AA199683_37	3700001	3810000
WP	AA199683_38	3800001	3910000
WP	AA199683_39	3900001	4010000
WP	AA199683_40	4000001	4110000
WP	AA199683_41	4100001	4210000
WP	AA199683_42	4200001	4310000
WP	AA199683_43	4300001	4403765

Query Match 100.0%; Score 21; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 3; Mismatches 0; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCGATGATGATTTACTCTCC 21

DB 85866 TCCGATGATGATTTACTCTCC 85886

RESULT 6

ACA47360/c

ACA47360 standard; DNA; 1275 BP.

AC ACA47360;

DT 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #29017.

KM Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

OS Staphylococcus haemolyticus.

XX WO200277183-A2.

XX 03-OCT-2002.
PD
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (BLIT-) BLITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
PI WPI; 2003-029926/02.
DR P-PSDB; ABU43490.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 35230; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC the gene product or that has an activity against a biological pathway
CC of the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1275 BP; 461 A; 179 C; 244 G; 391 T; 0 U; 0 Other;
SQ
Query Match 82.9%; Score 17.4; DB 8; Length 1275;
Best Local Similarity 94.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

XX 04-SEP-2000 (first entry)
DT
XX
XX DNA sequence whose expression is regulated by presence of opiates.
DE
XX
XX Opiate; haematopoietic cell; immune deficiency; blood protein disorder;
KW agammaglobulinemia; dysgammaglobulinemia; ataxia telangiectasia;
KW common variable immunodeficiency; DiGeorge Syndrome; HIV infection; HTLV;
KW Human T-cell leukaemia; lymphoma virus infection;
KW leucocyte adhesion deficiency syndrome; lymphopenia; hemoglobinuria;
KW phagocyte bactericidal dysfunction; severe combined immunodeficiency;
KW SCID; Wiskott-Aldrich Disorder; anemia; thrombocytopenia; ss.
XX
XX Mus musculus.
OS
XX
XX MO200028026-A2.
PN
XX
XX 18-MAY-2000.
PD
XX
XX 19-OCT-1999; 99WO-US024708.
PF
XX
XX 10-NOV-1998; 98US-0107806P.
PR
XX
XX (DIGI-) DIGITAL GENE TECHNOLOGIES INC.
PA
XX
XX Sutcliffe JG, Przewlocki R, Quan J, Thomas EA;
PI WPI; 2000-376541/32.
DR
XX
XX New polynucleotides whose expression is regulated by opiate
PT administration and withdrawal, useful for treating blood protein
PT disorders, ataxia telangiectasia and common variable immunodeficiency.
XX
XX Claim 3; Page 83; 85pp; English.
XX
XX AA15419-53 represent polynucleotides whose expression is regulated by
CC opiate administration and withdrawal. The polynucleotides are useful in
CC methods for preventing, treating or ameliorating a medical condition.
CC Diagnostic methods are used to detect mutations in the polynucleotides.
CC They could be used to increase differentiation and proliferation of
CC haematopoietic cells, including the pluripotent stem cells, in an effort
CC to treat those disorders associated with a decrease in certain (or many)
CC types haematopoietic cells. Examples of immunologic deficiency syndromes
CC include, but are not limited to blood protein disorders (e.g.
CC agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common
CC variable immunodeficiency, DiGeorge Syndrome, HIV infection, HTLV (Human
CC T-cell leukaemia/ lymphoma virus)-BLV infection, leucocyte adhesion
CC deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction,
CC severe combined immunodeficiency (SCID), Wiskott-Aldrich Disorder,
CC anemia, thrombocytopenia, or hemoglobinuria
XX
XX Sequence 397 BP; 100 A; 95 C; 81 G; 121 T; 0 U; 0 Other;
SQ
Query Match 81.0%; Score 17; DB 3; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 3 CGGATGATGATTTACTT 19
QY
DB 298 CGGATGATGATTTACTT 314

RESULT 8
ADF77343_03
Continuation (4 of 20) of ADF77343 from base 300001 (Lactic acid bacteria Lactobacillus
WP Sequence split into 20 fragments LOCUS ADF77343 Accession Adf77343
WP Fragment Name Begin End
WP ADF77343_00 1 11000
WP ADF77343_01 21000
WP ADF77343_02 31000
WP ADF77343_03 41000
WP ADF77343_04 51000
WP ADF77343_05 61000

WP ADF77343_06 600001 710000
WP ADF77343_07 700001 810000
WP ADF77343_08 800001 910000
WP ADF77343_09 900001 1010000
WP ADF77343_10 1000001 1110000
WP ADF77343_11 1100001 1210000
WP ADF77343_12 1200001 1310000
WP ADF77343_13 1300001 1410000
WP ADF77343_14 1400001 1510000
WP ADF77343_15 1500001 1610000
WP ADF77343_16 1600001 1710000
WP ADF77343_17 1700001 1810000
WP ADF77343_18 1800001 1910000
WP ADF77343_19 1900001 1983043

Query Match 81.0%; Score 17; DB 10; Length 110000;
Best Local Similarity 100.0%; Pred. No. 3,4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCCGGATGATGATTAC 17
DB 12164 TCCGGATGATGATTAC 12180

RESULT 9
ID ADQ21179 standard; DNA; 131239 BP.
XX ADQ21179;
AC ADQ21179;
XX 26-AUG-2004 (first entry)
DT
XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 3999.
DE
XX soft tissue sarcoma; cytosolic; gene therapy; vaccine; screening; human;
KW ds.
XX Homo sapiens.
OS
XX WO2004048938-A2.
FN
XX 10-JUN-2004.
PD
XX 26-NOV-2003; 2003WO-US038193.
PF
XX 26-NOV-2002; 2002US-0429739P.
PR
XX (PROT-) PROTEIN DESIGN LABS INC.
PA
XX Aziz N, Ginsburg WM, Zlotnick A;
PI WPI; 2004-441208/41.
DR
XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
XX Example 2; SEQ ID NO 3999; 210pp; English.
PS
XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual,
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytosolic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.

XX Sequence 131239 BP; 35649 A; 28535 C; 29022 G; 38033 T; 0 U; 0 Other;
SQ

Query Match 80.0%; Score 16.8; DB 12; Length 131239;
Best Local Similarity 90.0%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 CCGATGATGATTACTCC 21
DB 131217 CCGATGATGATTACTCC 131198

RESULT 10
ID AC192883 standard; DNA; 25 BP.
XX AC192883;
AC AC192883;
XX 14-OCT-2003 (first entry)
DT
XX Human microarray DNA oligonucleotide SEQ ID NO 92874.
DE
XX EST; ss; probe; expressed sequence tag; microarray; gene expression;
KW genetic variation; diallelic marker; polymorphism; human;
KW cross-species comparison.
XX
XX Homo sapiens.
OS
XX US2003104410-A1.
FN
XX 05-JUN-2003.
PD
XX 15-MAR-2002; 2002US-00098263.
PF
XX 16-MAR-2001; 2001US-0276759P.
PR
XX (AFPM-) AFPMETRIX INC.
PA
XX Miltmann MP;
PI WPI; 2003-567953/53.
DR
XX New array of nucleic acid probes, useful for in situ hybridization, in
PT Northern, Northern or dot-blot hybridization to identify or detect the
PT sequence or specific mutations of any gene.
PT
XX Claim 1; SEQ ID NO 92874; 9pp; English.
PS
XX The invention discloses a microarray comprising a plurality of nucleic
CC acid probes including one of 2,018,500 fully defined sequences, or its
CC perfect match, perfect mismatch, antisense match or antisense mismatch.
CC Also disclosed is a method of gene expression analysis. The array is used
CC in monitoring gene expression levels by hybridization to a DNA library,
CC in analysis of genetic variation or in hybridization of tag-labelled
CC compounds. The nucleic acid probes are specifically designed for analysis
CC of at least one target sequence. The method of analysis comprises
CC hybridizing at least one or more nucleic acids to at least two or more
CC nucleic acid probes and detecting the hybridization. The nucleic acid
CC probes are attached to a solid support. The analysis comprises monitoring
CC gene expression levels, identifying diallelic markers or polymorphisms,
CC or family members of a gene and a cross-species comparison. Each of the
CC nucleic acids further comprises a tag sequence. The array of nucleic acid
CC probes is useful in in situ hybridization, in Southern, Northern or dot-
CC blot hybridization to identify or detect the sequence or specific
CC mutations of any gene, in mapping the 5' termini of mRNA molecules by
CC primer extensions or in screening cDNA or genomic libraries or subclones
CC for additional subclones containing segments of DNA that have been
CC isolated and previously sequenced. The sequence presented is one of the
CC nucleic acid probes incorporated in the microarray. Note: The sequence
CC data for this patent can also be obtained in electronic format directly
CC from USPTO at seqdata.uspto.gov/sequence.html
XX
XX Sequence 25 BP; 7 A; 7 C; 8 G; 3 T; 0 U; 0 Other;

Query Match 77.1%; Score 16.2; DB 9; Length 25;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TCCGATGATGATTACCTCC 21
22 TCCGATGTTGAGGACTCC 2
Db
RESULT 11
AAC1729/c
ID AAC1729 standard; DNA; 268 BP.
XX
AC AAC1729;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 32922.
XX
XX Hybridisation assay; Genetic mapping; Gene expression control;
KM protein identification; signal transduction pathway; metabolic pathway;
KM promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS
PN BP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123546P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0128845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0135621P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.

PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 18-JUN-1999; 99US-0139817P.
PR 21-JUN-1999; 99US-0139899P.
PR 22-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140658P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142380P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144684P.
PR 20-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145132P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.

PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0148684P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0153363P.
PR 10-SEP-1999; 99US-0153707P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0158294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161924P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 77.1%; Score 16.2; DB 3; Length 288;
Best Local Similarity 85.7%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 TCCGGATGATGATTACCTCC 21
||||| ||||| |||||

DB 77 TCCGGTGTGATTACCTCC 57

RESULT 12
ADA58045
ADA58045 standard; cDNA; 304 BP.
AC ADA58045;
XX 20-NOV-2003 (first entry)
XX
XX Maize triose phosphate isomerase EST #199.
DE
XX
XX Plant; ss; EST; expressed sequence tag; transgenic; sucrose pathway;
KM triose phosphate isomerase; fructose 1, 6-bisphosphate aldolase;
KM fructose 1, 6-bisphosphate; fructose 6-phosphate 2-kinase;
KM phosphoglucosylisomerase; vacuolar hydrogen translocating-pyrophosphatase;
KM pyrophosphate-dependent fructolase fructose-6-phosphate phosphotransferase;
KM invertase; sucrose synthase; hexokinase; fructokinase;
KM nucleoside diphosphate kinase-kinase; NDP;
KM glucose-6-phosphate 1-dehydrogenase; phosphoglucumutase; UDP;
KM uridine diphosphate-glucose pyrophosphorylase; maize.
OS
XX Zea mays.
XX US2003135870-A1.
XX
XX 17-JUL-2003.
XX
XX 26-JAN-1999; 99US-00237183.
XX
XX 24-NOV-1997; 97US-0067000P.
PR 09-DEC-1997; 97US-0069472P.
PR 27-JAN-1998; 98US-0072888P.
PR 10-FEB-1998; 98US-0074201P.
PR 10-FEB-1998; 98US-0074280P.
PR 10-FEB-1998; 98US-0074281P.
PR 10-FEB-1998; 98US-0074282P.
PR 12-FEB-1998; 98US-0074565P.
PR 12-FEB-1998; 98US-0074566P.
PR 12-FEB-1998; 98US-0074567P.
PR 12-FEB-1998; 98US-0074789P.
PR 19-FEB-1998; 98US-0075459P.
PR 19-FEB-1998; 98US-0075460P.
PR 19-FEB-1998; 98US-0075461P.
PR 19-FEB-1998; 98US-0075462P.
PR 19-FEB-1998; 98US-0075463P.
PR 19-FEB-1998; 98US-0075464P.
PR 06-MAR-1998; 98US-0076912P.
PR 09-MAR-1998; 98US-0077229P.
PR 09-MAR-1998; 98US-0077230P.
PR 09-MAR-1998; 98US-0077231P.
PR 18-MAR-1998; 98US-0078368P.
PR 07-APR-1998; 98US-0080844P.
PR 27-APR-1998; 98US-0083067P.
PR 29-APR-1998; 98US-0083366P.
PR 29-APR-1998; 98US-0083367P.
PR 29-APR-1998; 98US-0083388P.
PR 29-APR-1998; 98US-0083389P.
PR 29-APR-1998; 98US-0083390P.
PR 13-MAY-1998; 98US-0085222P.
PR 13-MAY-1998; 98US-0085223P.
PR 13-MAY-1998; 98US-0085224P.
PR 21-MAY-1998; 98US-0086183P.
PR 21-MAY-1998; 98US-0086184P.
PR 21-MAY-1998; 98US-0086185P.
PR 21-MAY-1998; 98US-0086186P.
PR 21-MAY-1998; 98US-0086187P.
PR 01-JUN-1998; 98US-0087422P.
PR 16-JUN-1998; 98US-0089524P.
PR 18-JUN-1998; 98US-0089793P.
PR 18-JUN-1998; 98US-0089810P.

PR 18-JUN-1998; 98US-0089814P.
PR 22-JUN-1998; 98US-0090170P.
PR 26-JUN-1998; 98US-0090928P.
PR 29-JUN-1998; 98US-0091035P.
PR 30-JUN-1998; 98US-0091405P.
PR 08-JUL-1998; 98US-0092036P.
PR 09-SEP-1998; 98US-0099667P.
PR 09-SEP-1998; 98US-0099670P.
PR 09-SEP-1998; 98US-0099697P.
PR 16-SEP-1998; 98US-0100672P.
PR 16-SEP-1998; 98US-0100673P.
PR 16-SEP-1998; 98US-0100674P.
PR 16-SEP-1998; 98US-0100674P.
PR 21-SEP-1998; 98US-0101131P.
PR 21-SEP-1998; 98US-0101132P.
PR 22-SEP-1998; 98US-0101343P.
PR 22-SEP-1998; 98US-0101344P.
PR 22-SEP-1998; 98US-0101347P.
PR 22-SEP-1998; 98US-0101508P.
PR 25-SEP-1998; 98US-0101707P.
PR 13-OCT-1998; 98US-0104123P.
PR 13-OCT-1998; 98US-0104124P.
PR 13-OCT-1998; 98US-0104126P.
PR 13-OCT-1998; 98US-0104127P.
PR 13-OCT-1998; 98US-0104128P.
PR 18-NOV-1998; 98US-0108996P.
PR 19-NOV-1998; 98US-0109018P.
PR 24-NOV-1998; 98US-0019912P.
PR 08-DEC-1998; 98US-0021023P.
PR 11-DEC-1998; 98US-0111981P.
PR 22-DEC-1998; 98US-0113224P.
PR 12-JAN-1999; 99US-00229413.
XX
XX (CHEI/) CHEIKH N.
PA (FISH/) FISHER D K.
PA (LIU/) LIU J.
PI Chelkh N, Fisher DK, Liu J;
PI WPI; 2003-688722/65.
XX
XX
XX New maize or soybean enzymes and nucleic acid molecules associated with
PT the sucrose pathway, useful for genome mapping, gene identification and
PT analysis, plant breeding, or preparation of constructs for plant gene
PT expression.
XX
XX Claim 2; Page: 117pp; English.
XX
XX The invention relates to a substantially purified nucleic acid molecule
CC (appearing as AD857847 - ADA60660 that encodes a maize or soybean enzyme
CC or its fragment, associated with the sucrose pathway selected from:
CC triose phosphate isomerase, fructose 1,6-bisphosphate aldolase, fructose
CC 1,6-bisphosphate, fructose 6-phosphate 2-kinase, phosphoglucose isomerase,
CC vacuolar hydrotgen translocating-pyrophosphatase, pyrophosphate-dependent
CC fructose-6-phosphate phosphotransferase, invertase, sucrose synthase,
CC hexokinase, fructokinase, nucleoside diphosphate kinase (NDP)-kinase,
CC glucose-6-phosphate 1-dehydrogenase, phosphoglucomutase and uridine
CC diphosphate (UDP)-glucose pyrophosphorylase. Also included are a
CC substantially purified maize or soybean enzyme (or its fragment) and a
CC transformed plant having a nucleic acid molecule. Also disclosed are new
CC are purified antibodies capable of specifically binding to the maize or
CC soybean enzyme, determining a level or pattern of a plant sucrose pathway
CC enzyme in a plant cell or plant tissue, determining a mutation in a plant
CC whose presence is predictive of a mutation affecting the level or pattern
CC of a plant sucrose pathway enzyme, producing a plant containing an
CC overexpressed or reduced level of plant sucrose pathway enzyme, reducing
CC expression of a plant sucrose pathway enzyme in a plant and determining
CC an association between a polymorphism and a plant trait. The maize or
CC soybean enzymes and nucleic acid molecules are useful for genome mapping,
CC gene identification and analysis, plant breeding, or preparation of
CC constructs for plant gene expression and transgenic plants. The nucleic
CC acid molecules are also useful as markers or probes. The present sequence
CC is a maize BstI (expressed sequence tag) from a gene encoding a sucrose

CC pathway enzyme. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from USPTO at seqdata.uspto.gov/sequence.html?docID=20030135870.
XX
SQ Sequence 304 BP, 57 A, 78 C, 76 G, 93 T, 0 U, 0 Other;
Query Match 77.1%; Score 16.2; DB 9; Length 304;
Best Local Similarity 85.7%; Pred. No. 3,66+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 TCCGATGATGATTTACTCTCC 21
Db 219 TCCGATGCTGATTTCCCTCC 239
RESULT 13
ADP60670
ID ADP60670 standard; cDNA; 304 BP.
XX
AC ADP60670;
XX
DT 09-SEP-2004 (first entry)
XX
XX
DE Maize carbon assimilation pathway enzyme cDNA #1812.
XX
XX Carbon assimilation pathway enzyme; gene; ss; maize; corn;
KW ribulose-bisphosphate carboxylase; phosphoglycerate kinase;
KW glyceraldehyde 3-phosphate dehydrogenase; triose phosphate isomerase;
KW aldolase; fructose-1,6-bisphosphate; transketolase;
KW sedoheptulose-1,7-bisphosphatase; D-ribulose-5-phosphate-3-epimerase;
KW ribose-5-phosphate isomerase; ribose-5-phosphate kinase;
KW phosphoenolpyruvate carboxylase; NADP-dependent malate dehydrogenase;
KW aspartate aminotransferase; alanine aminotransferase;
KW NADP-dependent malic enzyme; NAD-dependent malic enzyme;
KW PEP carboxylase; pyruvate; phosphate dikinase; pyrophosphatase;
KW plant metabolic pathway; plant breeding.
XX
XX Zea mays.
OS
XX
XX US2004116682-A1.
XX
XX 17-JUN-2004.
XX
XX 16-NOV-2001; 2001US-00987899.
XX
XX 06-MAR-1998; 98US-0076912P.
PR 04-MAR-1999; 99US-00262979.
XX
XX (CHEI/) CHEIKH N.
PA (MILL/) MILLER P W.
PA (OCON/) O'CONNELL K M.
PA (LIU/) LIU J.
PI Chelkh N, Miller PW, O'connell KM, Liu J;
PI WPI; 2004-498291/47.
XX
XX New substantially purified nucleic acid molecule encoding a maize or
PT soybean carbon assimilation pathway enzyme, useful for gene mapping, gene
PT identification and analysis, plant breeding, and preparation of
PT constructs.
XX
XX Claim 2; SEQ ID NO 3739; 196pp; English.
XX
XX The invention relates to a substantially purified nucleic acid molecule
CC that encodes a maize or soybean carbon assimilation pathway enzyme or its
CC fragment. The maize or soybean carbon assimilation pathway enzyme or its
CC fragment is selected from ribulose-bisphosphate carboxylase,
CC phosphoglycerate kinase, glyceraldehyde 3-phosphate dehydrogenase, triose
CC phosphate isomerase, aldolase, fructose-1,6-bisphosphatase,
CC transketolase, sedoheptulose-1,7-bisphosphatase, D-ribulose-5-phosphate-3-
CC epimerase, ribose-5-phosphate isomerase, ribose-5-phosphate kinase,
CC phosphoenolpyruvate carboxylase, NADP-dependent malate dehydrogenase,

CC aspartate aminotransferase, alanine aminotransferase, NADP-dependent
CC malic enzyme, NAD-dependent malic enzyme, PEP carboxylase, pyruvate,
CC phosphate dikinase and pyrophosphatase. The invention also relates to a
CC substantially purified antibody or its fragment which is capable of
CC specifically binding to a specific maize or soybean carbon assimilation
CC pathway enzyme or its fragment, a transformed plant having a nucleic acid
CC molecule comprising an exogenous promoter region which functions in a
CC plant cell to cause the production of an mRNA molecule, and a method of
CC determining a level or pattern in a plant cell of a carbon assimilation
CC pathway enzyme in a plant metabolic pathway. The methods and compositions
CC of the invention are useful for gene mapping, gene identification and
CC analysis, plant breeding and preparation of constructs for use in plant
CC gene expression and transgenic plants. This sequence represents cDNA
CC encoding a maize carbon assimilation pathway enzyme of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html.

SQ Sequence 304 BP; 57 A; 78 C; 76 G; 93 T; 0 U; 0 Other;

Query Match 77.1%; Score 16.2; DB 12; Length 304;
Beet Local Similarity 85.7%; Pred. No. 3,6e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 TCCGGATGATGATTACCTCC 21
|||
Db 219 TCCGGATGCTGATTTCCTCC 239

RESULT 14

ADA57946
ADA57946 standard; cDNA; 316 BP.

AC ADA57946;

DT 20-NOV-2003 (First entry)

XX Maize triose phosphate isomerase EST #100.

XX Plant, ss; EST; expressed sequence tag; transgenic; sucrose pathway;
KM triose phosphate isomerase; fructose 1, 6-bisphosphate aldolase;
KM fructose 1, 6-bisphosphate; fructose 6-phosphate 2-kinase;
KM phosphoglucosyltransferase; vacuolar hydrogen translocating-pyrophosphatase;
KM pyrophosphatase-dependent fructose-6-phosphate phosphotransferase;
KM invertase; sucrose synthase; hexokinase; fructokinase;
KM nucleoside diphosphate kinase; NMP;
KM glucose-6-phosphate 1-dehydrogenase; phosphoglucosyltransferase; UDP;
KM uridine diphosphate-glucose pyrophosphorylase; maize.

OS Zea mays.

XX US2003135870-A1.

XX 17-JUL-2003.

PD 26-JAN-1999; 99US-00237183.

XX 24-NOV-1997; 97US-0067000P.

PR 09-DEC-1997; 97US-0069472P.

PR 27-JAN-1998; 98US-0072888P.

PR 10-FEB-1998; 98US-0074201P.

PR 10-FEB-1998; 98US-0074280P.

PR 10-FEB-1998; 98US-0074281P.

PR 10-FEB-1998; 98US-0074282P.

PR 12-FEB-1998; 98US-0074565P.

PR 12-FEB-1998; 98US-0074566P.

PR 12-FEB-1998; 98US-0074567P.

PR 19-FEB-1998; 98US-0074789P.

PR 19-FEB-1998; 98US-0075459P.

PR 19-FEB-1998; 98US-0075460P.

PR 19-FEB-1998; 98US-0075461P.

PR 19-FEB-1998; 98US-0075462P.

PR 19-FEB-1998; 98US-0075463P.

PR 19-FEB-1998; 98US-0075464P.

PR 06-MAR-1998; 98US-0076912P.

PR 09-MAR-1998; 98US-0077222P.

PR 09-MAR-1998; 98US-0077230P.

PR 09-MAR-1998; 98US-0077231P.

PR 18-MAR-1998; 98US-0078368P.

PR 07-APR-1998; 98US-0080844P.

PR 27-APR-1998; 98US-0083067P.

PR 29-APR-1998; 98US-0083386P.

PR 29-APR-1998; 98US-0083387P.

PR 29-APR-1998; 98US-0083388P.

PR 29-APR-1998; 98US-0083389P.

PR 29-APR-1998; 98US-0083390P.

PR 13-MAY-1998; 98US-0085223P.

PR 13-MAY-1998; 98US-0085224P.

PR 13-MAY-1998; 98US-0085225P.

PR 13-MAY-1998; 98US-0086183P.

PR 21-MAY-1998; 98US-0086184P.

PR 21-MAY-1998; 98US-0086185P.

PR 21-MAY-1998; 98US-0086186P.

PR 21-MAY-1998; 98US-0086187P.

PR 21-MAY-1998; 98US-0086188P.

PR 01-JUN-1998; 98US-0087422P.

PR 16-JUN-1998; 98US-0089524P.

PR 18-JUN-1998; 98US-0089793P.

PR 18-JUN-1998; 98US-0089810P.

PR 18-JUN-1998; 98US-0089811P.

PR 22-JUN-1998; 98US-0090170P.

PR 26-JUN-1998; 98US-0090928P.

PR 29-JUN-1998; 98US-0091035P.

PR 30-JUN-1998; 98US-0091405P.

PR 08-JUL-1998; 98US-0092036P.

PR 09-SEP-1998; 98US-0093667P.

PR 09-SEP-1998; 98US-0093670P.

PR 09-SEP-1998; 98US-0093671P.

PR 16-SEP-1998; 98US-0100672P.

PR 16-SEP-1998; 98US-0100673P.

PR 16-SEP-1998; 98US-0100674P.

PR 21-SEP-1998; 98US-0101130P.

PR 21-SEP-1998; 98US-0101131P.

PR 21-SEP-1998; 98US-0101132P.

PR 22-SEP-1998; 98US-0101343P.

PR 22-SEP-1998; 98US-0101344P.

PR 22-SEP-1998; 98US-0101347P.

PR 22-SEP-1998; 98US-0101508P.

PR 25-SEP-1998; 98US-0101707P.

PR 13-OCT-1998; 98US-0104123P.

PR 13-OCT-1998; 98US-0104124P.

PR 13-OCT-1998; 98US-0104126P.

PR 13-OCT-1998; 98US-0104128P.

PR 13-OCT-1998; 98US-0104129P.

PR 18-NOV-1998; 98US-0108996P.

PR 19-NOV-1998; 98US-0109018P.

PR 24-NOV-1998; 98US-00199129.

PR 08-DEC-1998; 98US-00210297.

PR 11-DEC-1998; 98US-0111981P.

PR 22-DEC-1998; 98US-0113224P.

PR 12-JAN-1999; 99US-00229413.

XX (CHEI//) CHEIKH N.
PA (FISH//) FISHER D K.
PA (LIU//) LIU J.

PI Cheikh N, Fisher DK, Liu J;

XX WPI; 2003-688722/65.

PT New maize or soybean enzymes and nucleic acid molecules associated with
PT the sucrose pathway, useful for genome mapping, gene identification and
PT analysis, plant breeding, or preparation of constructs for plant gene
PT expression.

PS Claim 2; Page; 117pp; English.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 19:19:37 ; Search time 71.266 Seconds
(without alignments)
482.163 Million cell updates/sec

Title: US-10-777-131A-1

Perfect score: 21

Sequence: 1 tcgcgatgatcttactctc 21

Scoring table: IDENTITY_NUC

Searched: Gapop 10.0 , Gapext 1.0
1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCUTS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	4403765	3 US-09-103-840A-2	Sequence 2, Appl1
2	21	100.0	4411529	3 US-09-103-840A-1	Sequence 1, Appl1
3	16.8	80.0	4308	4 US-09-614-221A-599	Sequence 599, App
4	16.2	77.1	1131	4 US-09-328-352-1902	Sequence 14902, A
5	16.2	77.1	44019	4 US-09-949-016-14902	Sequence 95, Appl
6	16	76.2	3510	4 US-09-265-585C-95	Sequence 12128, A
7	15.8	75.2	666	4 US-09-248-796A-12128	Sequence 181, App
8	15.4	73.3	415	4 US-09-220-132-181	Sequence 18981, A
9	15.4	73.3	601	4 US-09-949-016-18981	Sequence 18983, A
10	15.4	73.3	601	4 US-09-949-016-18982	Sequence 47061, A
11	15.4	73.3	601	4 US-09-949-016-18983	Sequence 47062, A
12	15.4	73.3	601	4 US-09-949-016-18984	Sequence 47063, A
13	15.4	73.3	601	4 US-09-949-016-18985	Sequence 2549, Ap
14	15.4	73.3	750	4 US-09-543-681A-2549	Sequence 131, App
15	15.4	73.3	1173	3 US-09-561-332-1	Sequence 777, App
16	15.4	73.3	1337	4 US-09-220-132-131	Sequence 13118, A
17	15.4	73.3	2742	4 US-09-799-451-777	Sequence 11809, A
18	15.4	73.3	107800	4 US-09-949-016-13118	Sequence 12201, A
19	15.4	73.3	116425	4 US-09-949-016-11809	Sequence 12201, A
20	15.4	73.3	227390	4 US-09-949-016-12201	Sequence 11305, A
21	15.4	73.3	227391	4 US-09-949-016-13365	Sequence 11308, A
22	15.4	73.3	636591	4 US-09-949-016-13388	Sequence 1211, Ap
23	15.2	72.4	346	4 US-09-513-999C-1211	Sequence 13712, A
24	15.2	72.4	358	4 US-09-621-976-13712	Sequence 2121, Ap
25	15.2	72.4	407	4 US-09-621-976-2121	

28	15.2	72.4	413	3 US-09-227-357-71	Sequence 71, Appl
29	15.2	72.4	427	4 US-09-513-999C-10820	Sequence 10820, A
30	15.2	72.4	601	4 US-09-949-016-60865	Sequence 60865, A
31	15.2	72.4	601	4 US-09-949-016-127401	Sequence 127401, A
32	15.2	72.4	2264	2 US-08-149-097D-26	Sequence 26, Appl
33	15.2	72.4	2404	4 US-09-269-717-16	Sequence 26, Appl
34	15.2	72.4	2634	3 US-08-984-386-26	Sequence 26, Appl
35	15.2	72.4	2634	3 US-08-450-562-26	Sequence 26, Appl
36	15.2	72.4	2634	3 US-08-984-709A-26	Sequence 26, Appl
37	15.2	72.4	2634	3 US-08-450-272-26	Sequence 26, Appl
38	15.2	72.4	2634	3 US-08-450-273-26	Sequence 26, Appl
39	15.2	72.4	2712	3 US-08-949-386-38	Sequence 38, Appl
40	15.2	72.4	2712	3 US-08-450-562-38	Sequence 38, Appl
41	15.2	72.4	2712	3 US-08-984-709A-38	Sequence 38, Appl
42	15.2	72.4	2712	3 US-08-450-272-38	Sequence 38, Appl
43	15.2	72.4	2909	3 US-08-450-273-38	Sequence 38, Appl
44	15.2	72.4	2909	3 US-08-104-158-1	Sequence 1, Appl
45	15.2	72.4	2909	4 US-09-609-040-1	Sequence 1, Appl

ALIGNMENTS

```
RESULT 1
US-09-103-840A-2
Sequence 2, 6294328
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103.840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2
Query Match 100.0%; Score 21; DB 3; Length 4403765;
Best local similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 1285866 TC CGCATGATGATTTACTCTC 1285866
Oy 1 TC CGCATGATGATTTACTCTC 21
DB 1285866 TC CGCATGATGATTTACTCTC 1285866
RESULT 2
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103.840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match          100.0%; Score 21; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCGATGATGATTACTCTCC 21
DB 1286397 TCCGATGATGATTACTCTCC 1286417

RESULT 3
US-09-614-221A-599
; Sequence 599, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
; APPLICANT: Karunamandaa, Balasubojini
; APPLICANT: Yu, Jaehyuk
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
; TITLE OF INVENTION: WITH STEROL SYNTHESIS AND METABOLISM
; FILE REFERENCE: 16516 075
; CURRENT APPLICATION NUMBER: US/09/614,221A
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 599
; LENGTH: 4308
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-599

Query Match          80.0%; Score 16.8; DB 4; Length 4308;
Best Local Similarity 90.0%; Pred. No. 70;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CCGATGATGATTACTCTCC 21
DB 1018 CCGAAGATGATTACTCCACC 1037

RESULT 4
US-09-328-352-1902
; Sequence 1902, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1902
; LENGTH: 1131
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1902

Query Match          77.1%; Score 16.2; DB 4; Length 1131;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCCGATGATGATTACTCTCC 21
DB 1098 TTCAGATGATGATTACTCCGCC 1118
```

```
RESULT 5
US-09-949-016-14902/C
; Sequence 14902, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14902
; LENGTH: 44019
; TYPE: DNA
; ORGANISM: Human
; FEATURES:
; NAME/KEY: misc.feature
; LOCATION: (1) - (44019)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14902

Query Match          77.1%; Score 16.2; DB 4; Length 44019;
Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCCGATGATGATTACTCTCC 21
DB 14173 TCCGATGATGATTACTCTCC 14153

RESULT 6
US-09-265-585C-95/C
; Sequence 95, Application US/09265585C
; Patent No. 6809234
; GENERAL INFORMATION:
; APPLICANT: Benfey, Phillip N.
; APPLICANT: Di Laurentio, Laura
; APPLICANT: Wypocka-Diller, Joanna
; APPLICANT: Malamy, Jocelyn E.
; APPLICANT: Pysh, Leonard
; APPLICANT: Helariutta, Yrjo
; APPLICANT: Bruce, Wesley
; APPLICANT: Lim, Jun
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses Thereof
; FILE REFERENCE: 5914-066
; CURRENT APPLICATION NUMBER: US/09/265,585C
; CURRENT FILING DATE: 1999-03-10
; PRIOR APPLICATION NUMBER: 08/842,445
; PRIOR FILING DATE: 1997-04-24
; PRIOR APPLICATION NUMBER: 08/638,617
; PRIOR FILING DATE: 1996-04-26
; NUMBER OF SEQ ID NOS: 152
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 95
; LENGTH: 3510
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (293)..(1855)
; NAME/KEY: CDS
; LOCATION: (2703)..(3143)
US-09-265-585C-95
```

Query Match 76.2%; Score 16; DB 4; Length 3510;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGGATGATGATTCTTA 16
 |||||
 Db 2386 TCCGGATGATGATTCTTA 2371

RESULT 7
 US-09-248-796A-12128
 ; Sequence 12128, Application US/09248796A
 ; Patent No. 6747137
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith Weinstein et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
 ; FILE REFERENCE: 107196.132
 ; CURRENT APPLICATION NUMBER: US/09/248,796A
 ; PRIOR FILING DATE: 1999-02-12
 ; PRIOR APPLICATION NUMBER: US 60/074,725
 ; PRIOR FILING DATE: 1998-02-13
 ; PRIOR APPLICATION NUMBER: US 60/096,409
 ; NUMBER OF SEQ ID NOS: 28208
 ; SEQ ID NO 12128
 ; LENGTH: 666
 ; TYPE: DNA
 ; ORGANISM: Candida albicans
 US-09-248-796A-12128

Query Match 75.2%; Score 15.8; DB 4; Length 666;
 Best Local Similarity 89.5%; Pred. No. 1.6e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCCGGATGATGATTCTT 19
 |||||
 Db 219 TCCGTGATGCTGATTCTT 237

RESULT 8
 US-09-220-132-181/c
 ; Sequence 181, Application US/09220132
 ; Patent No. 6506607
 ; GENERAL INFORMATION:
 ; APPLICANT: Shyjan, Andrew W.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
 ; FILE REFERENCE: 07334-074001
 ; CURRENT APPLICATION NUMBER: US/09/220,132
 ; PRIOR FILING DATE: 1998-12-23
 ; PRIOR APPLICATION NUMBER: US 60/079,303
 ; PRIOR FILING DATE: 1998-03-25
 ; PRIOR APPLICATION NUMBER: US 60/068,821
 ; PRIOR FILING DATE: 1997-12-24
 ; NUMBER OF SEQ ID NOS: 191
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 181
 ; LENGTH: 415
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(415)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-220-132-181

Query Match 73.3%; Score 15.4; DB 4; Length 415;
 Best Local Similarity 94.1%; Pred. No. 2.3e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 GATGATGATTTACTTCC 21

Db 283 GATGATGATTTACTTCC 267

RESULT 9
 US-09-949-016-18981/c
 ; Sequence 18981, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; FILE REFERENCE: CLO01307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; PRIOR FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 18981
 ; LENGTH: 601
 ; TYPE: DNA
 ; ORGANISM: Human
 US-09-949-016-18981

Query Match 73.3%; Score 15.4; DB 4; Length 601;
 Best Local Similarity 94.1%; Pred. No. 2.5e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CGGATGATGATTACTT 19
 |||||
 Db 389 CGGCTGATGATTACTT 373

RESULT 10
 US-09-949-016-18982/c
 ; Sequence 18982, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; FILE REFERENCE: CLO01307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; PRIOR FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 18982
 ; LENGTH: 601
 ; TYPE: DNA
 ; ORGANISM: Human
 US-09-949-016-18982

Query Match 73.3%; Score 15.4; DB 4; Length 601;
 Best Local Similarity 94.1%; Pred. No. 2.5e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CGGATGATGATTACTT 19
 |||||
 Db 394 CGGCTGATGATTACTT 378

RESULT 11

US-09-949-016-18983/c
; Sequence 18983, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18983
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-18983

Query Match 73.3%; Score 15.4; DB 4; Length 601;
Best Local Similarity 94.1%; Pred. No. 2.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CGGATGATGATTACT 19
|||
Db 486 CGGCTGATGATTACT 470

RESULT 12
US-09-949-016-47061/c
; Sequence 47061, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47061
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-47061

Query Match 73.3%; Score 15.4; DB 4; Length 601;
Best Local Similarity 94.1%; Pred. No. 2.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CGGATGATGATTACT 19
|||
Db 389 CGGCTGATGATTACT 373

RESULT 13
US-09-949-016-47062/c
; Sequence 47062, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;; FILE REFERENCE: CLO01307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; PRIOR FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 47062
;; LENGTH: 601
;; TYPE: DNA
;; ORGANISM: Human
US-09-949-016-47062

Query Match 73.3%; Score 15.4; DB 4; Length 601;
Best Local Similarity 94.1%; Pred. No. 2.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CGGATGATGATTACT 19
|||
Db 394 CGGCTGATGATTACT 378

RESULT 14
US-09-949-016-47063/c
; Sequence 47063, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47063
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-47063

Query Match 73.3%; Score 15.4; DB 4; Length 601;
Best Local Similarity 94.1%; Pred. No. 2.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CGGATGATGATTACT 19
|||
Db 486 CGGCTGATGATTACT 470

RESULT 15
US-09-543-681A-2549
; Sequence 2549, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 2549
 ; LENGTH: 750
 ; TYPE: DNA
 ; ORGANISM: Proteus mirabilis
 US-09-543-681A-2549

Query Match 73.3%; Score 15.4; DB 4; Length 750;
 Best Local Similarity 94.1%; Pred. No. 2.6e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GGATGATGATTACCTC 20
 |||||
 Db 57 GGATGATGATTACCTC 73

Search completed: November 6, 2005, 23:09:34
 Job time : 92.266 secs

This Page Blank (user)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 20:56:42 ; Search time 523.66 Seconds
(without alignments)
331.640 Million cell updates/sec

Title: US-10-777-131A-1

Perfect score: 21

Sequence: 1 tccgcgatgcattactctcc 21

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 9794790 seqs, 413490567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
- 22: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
- 23: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 24: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUBCOMB.seq.*
- 25: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
- 26: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 27: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 28: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	US-10-216-817-1	Sequence 1, Appl
2	21	100.0	21	US-10-777-131A-1	Sequence 1, Appl
3	21	100.0	306	US-10-755-415-92	Sequence 92, Appl
4	21	100.0	823	US-10-216-817-19	Sequence 19, Appl
5	21	100.0	823	US-10-777-131A-19	Sequence 19, Appl

C	6	17.8	84.8	669	21	US-10-425-115-10736	Sequence 10736, A
	7	17.8	84.8	4509	20	US-10-437-963-22127	Sequence 22127, A
	8	17.8	84.8	4596	20	US-10-437-963-22129	Sequence 22129, A
C	9	17.4	82.9	340	19	US-10-424-599-89548	Sequence 89548, A
C	10	17.4	82.9	403	19	US-10-424-599-87436	Sequence 87436, A
C	11	17.4	82.9	1275	18	US-10-282-122A-35230	Sequence 35230, A
C	12	16.8	80.0	397	20	US-10-437-963-63692	Sequence 63692, A
	13	16.8	80.0	1821	20	US-10-437-963-52292	Sequence 52292, A
	14	16.8	80.0	2181	20	US-10-437-963-52293	Sequence 52293, A
	15	16.8	80.0	4308	20	US-10-793-639-599	Sequence 599, A
C	16	16.8	80.0	131239	21	US-10-723-860-3999	Sequence 3999, A
C	17	16.8	80.0	131239	24	US-10-756-149-3963	Sequence 3963, A
	18	16.4	78.1	581	14	US-10-027-632-286839	Sequence 286839, A
	19	16.4	78.1	581	18	US-10-437-963-286839	Sequence 286839, A
	20	16.4	78.1	2250	20	US-10-437-963-100004	Sequence 100004, A
C	21	16.2	77.1	25	16	US-10-098-2638-92874	Sequence 92874, A
	22	16.2	77.1	249	21	US-10-425-115-84686	Sequence 84686, A
	23	16.2	77.1	304	10	US-09-237-183A-199	Sequence 199, A
	24	16.2	77.1	304	11	US-09-237-183A-197	Sequence 197, A
	25	16.2	77.1	316	10	US-09-237-183A-100	Sequence 100, A
	26	16.2	77.1	316	11	US-09-237-183A-101	Sequence 101, A
	27	16.2	77.1	325	10	US-09-237-183A-101	Sequence 101, A
	28	16.2	77.1	325	11	US-09-237-183A-101	Sequence 101, A
	29	16.2	77.1	393	19	US-10-425-114-2846	Sequence 2846, A
	30	16.2	77.1	398	10	US-09-237-183A-197	Sequence 197, A
	31	16.2	77.1	398	11	US-09-237-183A-197	Sequence 197, A
	32	16.2	77.1	483	13	US-09-925-065A-219459	Sequence 219459, A
	33	16.2	77.1	483	13	US-09-925-065A-219460	Sequence 219460, A
C	34	16.2	77.1	582	21	US-10-425-115-30270	Sequence 30270, A
C	35	16.2	77.1	592	24	US-10-972-079-87898	Sequence 87898, A
	36	16.2	77.1	600	24	US-10-972-079-87897	Sequence 87897, A
	37	16.2	77.1	645	14	US-10-027-632-280976	Sequence 280976, A
	38	16.2	77.1	645	18	US-10-027-632-280976	Sequence 280976, A
	39	16.2	77.1	656	14	US-10-027-632-212304	Sequence 212304, A
	40	16.2	77.1	656	14	US-10-027-632-212305	Sequence 212305, A
	41	16.2	77.1	656	14	US-10-027-632-212306	Sequence 212306, A
	42	16.2	77.1	656	18	US-10-027-632-212306	Sequence 212306, A
	43	16.2	77.1	656	18	US-10-027-632-212305	Sequence 212305, A
	44	16.2	77.1	656	18	US-10-027-632-212306	Sequence 212306, A
	45	16.2	77.1	726	19	US-10-425-114-6761	Sequence 6761, A

ALIGNMENTS

RESULT 1
US-10-216-817-1
; Sequence 1, Application US/10216817
; Publication No. US20030129619A1
GENERAL INFORMATION:
APPLICANT: GICQUEL, BRIGITTE
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
FILE REFERENCE: 03495.0233-00000
CURRENT APPLICATION NUMBER: US/10/216,817
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: 60/311,824
PRIOR FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: 60/313,523
PRIOR FILING DATE: 2001-08-21
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 21
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Primer
US-10-216-817-1
Query Match 100.0%; Score 21; DB 16; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGATGATGATTACTCC 21
Db 1 TCCGATGATGATTACTCC 21

RESULT 2
US-10-777-131A-1

; Sequence 1, Application US/10777131A
; Publication No. US2005026216A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; TITLE OF INVENTION: IN GENES OF THE mult FAMILY
; FILE REFERENCE: 02356.0090-00000
; CURRENT FILING DATE: US/10/777,131A
; PRIOR APPLICATION NUMBER: 2004-02-13
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 1
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
US-10-777-131A-1

Query Match 100.0%; Score 21; DB 22; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGATGATGATTACTCC 21
Db 1 TCCGATGATGATTACTCC 21

RESULT 3

US-10-755-415-92/C
; Sequence 92, Application US/10755415
; Publication No. US20050136480A1
; GENERAL INFORMATION:
; APPLICANT: BRAHMACHARI, SAMIR KUMAR
; APPLICANT: DASH, DEBASIS
; APPLICANT: SHARMA, RAMAKANT
; APPLICANT: MAHESHWARI, JITENDRA KUMAR
; TITLE OF INVENTION: A COMPUTER BASED VERSATILE METHOD FOR IDENTIFYING PROTEIN CODING
; TITLE OF INVENTION: DNA SEQUENCES USEFUL AS DRUG TARGETS
; FILE REFERENCE: 026033-00029
; CURRENT FILING DATE: US/10/755,415
; PRIOR APPLICATION NUMBER: 2004-01-13
; PRIOR FILING DATE: 2003-12-05
; NUMBER OF SEQ ID NOS: 373
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 92
; LENGTH: 306
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-755-415-92

Query Match 100.0%; Score 21; DB 24; Length 306;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGATGATGATTACTCC 21

Db 191 TCCGATGATGATTACTCC 171

RESULT 4
US-10-216-817-19

; Sequence 19, Application US/10216817
; Publication No. US20030129619A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; FILE REFERENCE: 03495.0233-00000
; CURRENT FILING DATE: US/10/216,817
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 823
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-216-817-19

Query Match 100.0%; Score 21; DB 16; Length 823;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGATGATGATTACTCC 21
Db 5 TCCGATGATGATTACTCC 25

RESULT 5
US-10-777-131A-19

; Sequence 19, Application US/10777131A
; Publication No. US2005026216A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; FILE REFERENCE: 02356.0090-00000
; CURRENT FILING DATE: US/10/777,131A
; PRIOR APPLICATION NUMBER: 2004-02-13
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 19
; LENGTH: 823
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-777-131A-19

Query Match 100.0%; Score 21; DB 22; Length 823;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGATGATGATTACTCC 21
Db 5 TCCGATGATGATTACTCC 25

RESULT 6

```

US-10-425-115-10736/c
: Sequence 10736, Application US/10425115
: Publication No. US20040214272A1
: GENERAL INFORMATION:
: APPLICANT: La Rosa, Thomas J.
: APPLICANT: Kovallik, David K.
: APPLICANT: Zhou, Yihua
: APPLICANT: Cao, Yongwei
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
: TITLE OF INVENTION: Plants
: FILE REFERENCE: 38-21(53222)B
: CURRENT APPLICATION NUMBER: US/10/425,115
: CURRENT FILING DATE: 2003-04-28
: NUMBER OF SEQ ID NOS: 369326
: SEQ ID NO 10736
: LENGTH: 669
: TYPE: DNA
: ORGANISM: Zea mays
: FEATURE:
: OTHER INFORMATION: Clone ID: MRT4577_109792C.1
: US-10-425-115-10736

```

Query Match	84.8%	Score 17.8;	DB 21;	Length 669;
Best Local Similarity	90.5%;	Pred. No. 1.4e+02;		
Matches 19; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

Qy 1 TCCGATGATGATTACCTCC 21
Db 566 TCCGATGATGATTTCCTCC 546

```

RESULT 7
US-10-437-963-22127
: Sequence 22127, Application US/10437963
: Publication NO. US20040123343A1
: GENERAL INFORMATION:
: APPLICANT: La Rosa, Thomas J.
: APPLICANT: Kovalic, David K.
: APPLICANT: Zhou, Yihua
: APPLICANT: Cao, Yongwei
: APPLICANT: Wu, Wei
: APPLICANT: Boukharov, Andrey A.
: APPLICANT: Barbazuk, Brad
: APPLICANT: Li, Ping
: TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
: TITLE OF INVENTION: Plant and Uses Thereof for Plant Improvement
: FILE REFERENCE: 38-21(53221) B
: CURRENT APPLICATION NUMBER: US/10/437, 963
: CURRENT FILING DATE: 2003-05-14
: NUMBER OF SEQ ID NOS: 204966
: SEQ ID NO 22127
: LENGTH: 4509
: TYPE: DNA
: ORGANISM: Oryza sativa
: FEATURE:
: OTHER INFORMATION: Clone ID: PAT_MRT4530_27332C.1
US-10-437-963-22127

```

Query Match	84.8%	Score 17.8;	DB 20;	Length 4509;
Best Local Similarity	90.5%	Pred. No. 2e+02;		
Matches 19; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

```
QY      1 TCCGATGATGATTACCTC 21
        |||||
Db     165 TCCGATGAGATTACCACC 185
```

RESULT 8
US-10-437-963-22129
; Sequence 22129, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.

```

1  APPLICANT: Kovalic, David K.
2  APPLICANT: Zhou, Yihua
3  APPLICANT: Cao, Yongwei
4  APPLICANT: Mu, Wei
5  APPLICANT: Boultharov, Andrey A.
6  APPLICANT: Barabzik, Brad
7  APPLICANT: Li, Ping
8  TITLE OF INVENTION: Rice Nucleic Acid Molecules Associated With
9  TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
10 FILE REFERENCE: 38-21(53221)B
11 CURRENT APPLICATION NUMBER: US/10/437,963
12 CURRENT FILING DATE: 2003-05-14
13 NUMBER OF SEQ ID NOS: 204966
14 SEQ ID NO 22129
15 LENGTH: 4596
16 TYPE: DNA
17 ORGANISM: Oryza sativa
18 FEATURE:
19 OTHER INFORMATION: Clone ID: PAT_MRT4530_27334C.1
20 US-10-437-963-22129

```

Query Match	84.8%	Score 17.8	DB 20	Length 4596
Best Local Similarity	90.5%	Pred. No. 2e+02		
Matches 19; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

Dy 1 TCCGATGATTTACTCC 21
|||
Db 165 TCCGATGAAGATTACACC 185

```

RESULT 9
US-10-424-599-89548/c
; Sequence: 89548, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 89548
; LENGTH: 340
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_51872C.1
US-10-424-599-89548

```

Query Match	82.9%	Score 17.4;	DB 19;	Length 340;
Best Local Similarity	94.7%	Pred. No. 1.9e+02;		
Matches 18; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY		2	CCGATGATGATTACCTC	20
Dd		181	CCGATGATGATTACCCC	163

RESULT 10
 US-10-424-599-87436/C
 ; Sequence 87436, Application US/10424599
 ; Publication No. US20040031072A1
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: La Rosa Thomas J
 ;
 ; APPLICANT: Kovalic David K
 ;
 ; APPLICANT: Zhou Yihua
 ;
 ; APPLICANT: Cao Yongwei
 ;
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

```
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 87436
LENGTH: 403
TYPE: DNA
ORGANISM: glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_49965C.1
US-10-424-599-87436

Query Match      82.9%; Score 17.4; DB 19; Length 403;
Best Local Similarity 94.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      2 CCGATGATGATTTACCTC 20
Db      235 CCGATGATGATTTACCCC 217

RESULT 11
US-10-282-122A-35230/c
Sequence 35230, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haeselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 35230
LENGTH: 1275
TYPE: DNA
ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-35230

Query Match      82.9%; Score 17.4; DB 18; Length 1275;
Best Local Similarity 94.7%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 TCCGATGATGATTTACCT 19
Db      360 TTCGATGATGATTTACCT 342

RESULT 12
US-10-437-963-63692/c
Sequence 63692, Application US/10437963
Publication No. US2004012343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 63692
LENGTH: 397
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_6490C.1
US-10-437-963-63692

Query Match      80.0%; Score 16.8; DB 20; Length 397;
Best Local Similarity 90.0%; Pred. No. 4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      2 CCGATGATGATTTACCTCC 21
Db      40 CCGATGATGATTTACCCCC 21

RESULT 13
US-10-437-963-52292
Sequence 52292, Application US/10437963
Publication No. US2004012343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 52292
LENGTH: 1821
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_54601C.1
US-10-437-963-52292

Query Match      80.0%; Score 16.8; DB 20; Length 1821;
Best Local Similarity 90.0%; Pred. No. 5.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Oy 2 CCGATGATGATTACTCC 21
 |||||
 Db 1207 CCTGAAGATGATTACTCC 1226

Search completed: November 7, 2005, 05:20:32
 Job time : 525.66 secs

RESULT 14
 US-10-437-963-52293
 ; Sequence 52293, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Bardazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 52293
 ; LENGTH: 2181
 ; TYPE: DNA
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_54602C.1
 US-10-437-963-52293

Query Match 80.0%; Score 16.8; DB 20; Length 2181;
 Best Local Similarity 90.0%; Pred. No. 5.4e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 CCGATGATGATTACTCC 21
 |||||
 Db 1567 CCTGAAGATGATTACTCC 1586

RESULT 15
 US-10-793-639-599
 ; Sequence 599, Application US/10793639
 ; Publication No. US20040199940A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Karunanandaa, Balasubramanian
 ; APPLICANT: Yu, Jaehyuk
 ; APPLICANT: Kishore, Ganesh M.
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
 ; TITLE OF INVENTION: WITH STEROL SYNTHESIS AND METABOLISM
 ; FILE REFERENCE: 16516.075
 ; CURRENT APPLICATION NUMBER: US/10/793,639
 ; CURRENT FILING DATE: 2004-03-05
 ; PRIOR APPLICATION NUMBER: US/09/614,221A
 ; PRIOR FILING DATE: 2000-07-11
 ; PRIOR APPLICATION NUMBER: US 60/142,981
 ; PRIOR FILING DATE: 1999-07-12
 ; NUMBER OF SEQ ID NOS: 626
 ; SEQ ID NO 599
 ; LENGTH: 4308
 ; TYPE: DNA
 ; ORGANISM: Saccharomyces cerevisiae
 US-10-793-639-599

Query Match 80.0%; Score 16.8; DB 20; Length 4308;
 Best Local Similarity 90.0%; Pred. No. 6.2e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 CCGATGATGATTACTCC 21
 |||||
 Db 1018 CCGAAGATGATTACTCC 1037

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 02:38:52 ; Search time 2100.89 Seconds
(without alignments)
380.481 Million cell updates/sec

Title: US-10-777-131A-1

Perfect score: 21

Sequence: 1 tcggatgatgatctaccctcc 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hic: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gsa1: *
9: gb_gsa2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	90.5	773	BZ499562	BZ499562 BONMG24TR
2	18.4	87.6	668	BP136446	BP136446 BP136446
3	18	85.7	764	CP439390	CP439390 EST675735
4	17.8	84.8	948	CL965683	CL965683 OSIFCC012
5	17.8	84.8	983	BM472190	BM472190 AGENCOURT
6	17.4	82.9	200	AV340408	AV340408 AV340408
7	17.4	82.9	805	CO809020	CO809020 AGENCOURT
8	17	81.0	196	CA328214	CA328214 UI-M-FIO-
9	17	81.0	289	AA254497	AA254497 val1c04.r
10	17	81.0	286	AV164857	AV164857 AV164857
C 11	17	81.0	320	BG060357	BG060357 L0900H05-
C 12	17	81.0	326	BE634382	BE634382 uv67B02.Y
C 13	17	81.0	337	BX525225	BX525225 BX525225
C 14	17	81.0	399	AM046909	AM046909 UI-M-BHI-
C 15	17	81.0	401	BY609421	BY609421 BY609421
C 16	17	81.0	414	BE991046	BE991046 UI-M-BZ1-
C 17	17	81.0	414	BE991225	BE991225 UI-M-BZ1-
C 18	17	81.0	416	BX636140	BX636140 BX636140
C 19	17	81.0	417	AA085413	AA085413 zn07906.s
20	17	81.0	418	CF744036	CF744036 UI-M-GVO-
21	17	81.0	425	BU611514	BU611514 UI-M-FIO-
22	17	81.0	466	BF582889	BF582889 602101336
23	17	81.0	472	BG868769	BG868769 602787607
C 24	17	81.0	484	AA880291	AA880291 vx39607.r

C 25	17	81.0	524	4	BM241795	BM241795 K0622E11-
C 26	17	81.0	538	7	BR853277	BR853277 uw87C03.X
C 27	17	81.0	577	2	CK616236	CK616236 ou14e11.Y
C 28	17	81.0	623	2	AM550920	AM550920 L0071B10-
29	17	81.0	649	1	AI152769	AI152769 udl17h10.r
30	17	81.0	722	1	AF122567	AF122567 AF122567
C 31	17	81.0	857	8	CC435783	CC435783 PUEB043TD
C 32	16.8	80.0	193	2	BR461953	BR461953 BR461953
33	16.8	80.0	256	9	CL221696	CL221696 ZMMB0C052
34	16.8	80.0	263	9	CR097239	CR097239 Reverse 8
C 35	16.8	80.0	394	2	BR432151	BR432151 EST198680
C 36	16.8	80.0	462	2	BR435941	BR435941 EST406935
C 37	16.8	80.0	486	2	BR460443	BR460443 EST411662
C 38	16.8	80.0	486	4	BM405690	BM405690 EST580017
C 39	16.8	80.0	500	2	BF051758	BF051758 EST1437005
C 40	16.8	80.0	507	2	BR435744	BR435744 EST406822
C 41	16.8	80.0	523	9	CL326924	CL326924 RPI144.24
C 42	16.8	80.0	527	4	BM535572	BM535572 EST588594
C 43	16.8	80.0	538	2	BR460424	BR460424 EST411643
44	16.8	80.0	541	5	BX332222	BX332222 BX332222
45	16.8	80.0	593	5	BP505841	BP505841 BP505841

ALIGNMENTS

RESULT 1
BZ499562/c
LOCUS
DEFINITION BONMG24TR BO.1.6.2 KB tot Brassica oleracea genomic clone BONMG24,
genomic survey sequence.
ACCESSION BZ499562
VERSION BZ499562.1 GI:27014783
KEYWORDS
SOURCE
ORGANISM Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosid II; Brassicales; Brassicaceae; Brassica.
REFERENCE
AUTHORS Town,C.D., Van Aken,S., Uterback,T., Koo,H. and Frazer,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES

source
1..773
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO100DH3"
/db_xref="taxon:3712"
/clone="BONMG24"
/clone_1lb="BO.1.6.2 KB tot"
/note="Vector: pHOSt1 site 1; BstXI; 1.6-2 kb sheared
total DNA inserted into pHOSt1 using BstXI linkers"

ORIGIN

Query Match 90.5%; Score 19; DB 8; Length 773;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCGATGATGATTTACTTC 20
|||||
DB 248 CCGATGATGATTTACTTC 230

```

RESULT 2
LOCUS      Bp136446                      668 bp    mRNA    linear    EST 17-JUL-2003
DEFINITION Bp136446 MAT001 Nicotiana tabacum cDNA clone BY9044, mRNA sequence.
ACCESSION  Bp136446
VERSION     Bp136446.1 GI:32890036
KEYWORDS   EST.
SOURCE      Nicotiana tabacum (common tobacco)
ORGANISM   Nicotiana tabacum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asterids; Lamiales; Solanales; Solanaceae; Nicotiana.
REFERENCE  1 (bases 1 to 668)
            Matsunaka, K., Tashiro, G., Horiuchi, T., Demura, T. and Fukuda, H.
            Profiling growth-phase dependent gene expression of tobacco BY-2
            cells by comprehensive microarray analysis
            Unpublished (2003)
JOURNAL    Contact: Ken Matsunaka
            Morphogenesis Research Group
            RIKEN Plant Science Center
            1-7-2 Suenohirocho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9575
            Fax: 81-45-503-9573
            Email: by2@psc.riken.go.jp URL: http://mrg.psc.riken.go.jp/stc/
            The cDNA library was constructed from mRNA isolated from lag (9 h),
            lag (72 h) and stationary (7 days) old BY-2 cells.
FEATURES   Location/Qualifiers
            source          1..668
                        /organism="Nicotiana tabacum"
                        /mol_type="mRNA"
                        /cultiyar="Bright Yellow No.2"
                        /db_xref="taxon:4097"
                        /clone="BY9044"
                        /cell_line="BY-2"
                        /note="Vector: pGEM-T easy; primer: M13 forward; mRNA
                        obtained from lag, lag and stationary phase cells"
ORIGIN
Query Match      87.6%; Score 18.4; DB 5; Length 668;
Best Local Similarity 95.0%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 CCGATGATGATTACTCTCC 21
    |||||
Db 298 CCGATGATGATTACTCTCC 317

RESULT 3
LOCUS      CF439390                      764 bp    mRNA    linear    EST 04-SEP-2003
DEFINITION EST675735 normalized cDNA library of onion Allium cepa cDNA clone
ACCESSION  CF439390
VERSION     CF439390.1 GI:34462080
KEYWORDS   EST.
SOURCE      Allium cepa (onion)
ORGANISM   Allium cepa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;
            Allium.
REFERENCE  1 (bases 1 to 764)
            Havery, M.J., Cheung, F., Van Aken, S., Uteback, T. and Town, C.D.
            Expressed Sequence Tags from a normalized library of mixed onion
            tissues (Allium cepa)
            Unpublished (2003)
JOURNAL    Contact: Havery MW
            Department of Horticulture
            USDA-ARS and University of Wisconsin
            1575 Linden Drive, Madison, WI 53706, USA
            Tel: 608-262-1830
            Fax: 608-262-4743
            Email: mjhavery@facstaff.wisc.edu
COMMENT

```

```

TIGR sequence name ACAX777R. For more information:
http://haverlab.hort.wisc.edu
Seq primer: CAG GAA ACA GCT ATG ACC.
Location/Qualifiers
1..764
/organism="Allium cepa"
/mol_type="mRNA"
/cultivar="Red Creole(bulbs), unknown(callus), Ebano &
Texas legend(rosetts)"
/db_xref="taxon:4679"
/clone="ACAX77"
/tissue_type="Callus, roots, and young bulbs"
/clone_lib="normalized cDNA library of onion"
/note="Vector: pCMVSPORT6.1-ccdb (Invitrogen); Site 1:
EcoRV (5'); Site 2: NotI (3'); Equal molar amounts of mRNA
from callus, roots, and young bulbs were combined to
synthesize the library. Normalization to enrich for
low-copy transcripts was performed by proprietary
techniques of Invitrogen."
ORIGIN
Query Match      85.7%; Score 18; DB 7; Length 764;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4 GGATGATGATTACTCTCC 21
    |||||
Db 622 GGATGATGATTACTCTCC 639

RESULT 4
LOCUS      CL965683                      948 bp    DNA    linear    GSS 21-SEP-2004
DEFINITION OsIFCC012507 Oryza sativa Expressed library Oryza sativa (indica
cultivar-group) genomic, genomic survey sequence.
ACCESSION  CL965683
VERSION     CL965683.1 GI:52386049
KEYWORDS   GSS.
SOURCE      Oryza sativa (indica cultivar-group)
ORGANISM   Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE  1 (bases 1 to 948)
            Ma, L., Mang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
            Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
            Wong, G. K. S., Deng, X. W. and Wang, J.
            An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
            Unpublished (2004)
JOURNAL    Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES   Location/Qualifiers
            source          1..948
                        /organism="Oryza sativa (indica cultivar-group)"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:39946"
                        /clone_lib="Oryza sativa Expressed library"
                        /note="Oryza sativa exon trapped genomic sequences"
ORIGIN
Query Match      84.8%; Score 17.8; DB 9; Length 948;
Best Local Similarity 90.5%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 TCCGATGATGATTACTCTCC 21

```

[illegible]

TITLE
JOURNAL
COMMENT

Meisuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, P., Tateo, M., Tomioka, N., Tanaka, Y., Watanabe, S., Yamamori, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. RIKEN Mouse ESTs (Kono, H., et al. 1999) Unpublished (1999)

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/
Sasaki, N., Izawa, M., Wataniki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh, M., Kikunishi, T., Akiyama, D., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

Location/Qualifiers

1. 200
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6430514N02"
/sex="male"
/tissue_type="olfactory brain"
/dev_stage="adult"
/lab_note="DH10B"
/clone_id="RIKEN full-length enriched, adult male olfactory bulb"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15'
GAGAGAGAGAGATCCAGAGCTTTTTTTTTTTTTTNN 3'. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequence enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGAGAGAGATCCAGCTTAATTAATTAATCCCCCCCC 3'. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified plasmid pUC19(+) after bulk excision from Lambda P/LC I. Cloning sites, 5' end: Sali; 3' end: BamHI"

ORIGIN

Query Match 82.9%; Score 17.4; DB 1; Length 200;
Best Local Similarity 94.7%; Prid. No. 4,4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 TCGGATGATGATTTTACT 19
|||||
Db 100 TCGGATGATGATTTTACT 118

RESULT 7	COB09020	805 bp	mRNA	linear	EST 06-AUG-2004
LOCUS	COB09020				
DEFINITION	AGNCOURT_30341687 NIH_ZGC_5 Danio rerio cDNA clone IMAGE:7396239				
ACCESSION	COB09020				
VERSION	COB09020.1	GI:51026281			
KEYWORDS	EST.				
SOURCE	Danio rerio (zebrafish)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes; Cyprinidae; Danio.				
REFERENCE	1 (bases 1 to 805)				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: cgapbs-remail.nih.gov Tissue Procurement: John Postlethwait, University of Oregon cDNA Library Preparation: Dr. Sumio Sugano cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Plate: LLM15579 row: 3 column: 13 High quality sequence scop: 501. Location/Qualifiers 1..805 /organism="Danio rerio" /mol_type="mRNA" /db_xref="taxon:7955" /clone_image="7396239" /tissue_type="contains eggs from one female, from all stages of development as well as support cells" /lab_host="DH10B Tona" /clone_lib="NIH ZGC 5" /note="Organ: ovary; Vector: pME18-FU3; Site 1: DraIII; Site 2: DraIII; 1st strand cDNA was primed with an oligo(dT) primer [GGCGCTGAGACGGCTCATGTGCGCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [GGCCUACUGG], digested and directionally cloned into distinct DraIII sites of the pME18-FU3. Library size selected for 1.0 kb, with a average insert size of ~1.2kb Library constructed by Yutaka Suzuki (University of Tokyo Institute of Medical Science). Custom primers recommended for sequencing: 5' end primer 5'-GGATGTGCTTTACTTCA-3' and 3' end primer 5'-CGACTGCAGCTCAGACCA-3'. Note: This is a Zebrafish Gene Collection (ZGC) library"				
ORIGIN					
Query Match	82.9%;	Score 17.4;	DB 7;	Length 805;	
Best Local Similarity	94.7%;	Pred. No. 5.5e+02;			
Matches	18;	Conservative 0;	Indels 1;	Gaps 0;	
Qy	1	TCCGGATGATGATTACTT	19		
Db	750	TCCGGAAGATGATTACTT	768		
RESULT 8	CA328214	196 bp	mRNA	linear	EST 09-JUL-2003
LOCUS	CA328214				
DEFINITION	UI-M-FPO-cdb-k-07-0-UI-r1 NIH BMAP_FPO Mus musculus cDNA clone				
ACCESSION	CA328214				
VERSION	CA328214.1	GI:24546312			
KEYWORDS	EST.				

```

SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
REFERENCE   1 (bases 1 to 196)
AUTHORS     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE       NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL     National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D. Email: cgaabs-remail.nih.gov Tissue Procurement: Dr. Jim Lin, University of Iowa CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLB at: http://image.lnl.gov This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES
source      Location/Qualifiers
            1..196
               /organism="Mus musculus"
               /mol_type="mRNA"
               /strain="C57BL/6"
               /db_xref="taxon:10090"
               /clone="IMAGE:6827360"
               /issue_type="whole brain"
               /dev_stage="embryo 13.5,14.5,16.5,17.5dpoc"
               /lab_host="MDH10B (T1 phage resistant)"
               /clone_lib="NIH-BMAP FYO"
               /note="Organ: Brain; Vector: pYX-Asc; Site:1: EcoR I; Site 2: Not I; The library was constructed according Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NciI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the university Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemmi Chin, Ph.D., program coordinator."

ORIGIN
Query Match      81.0%; Score 17; DB 6; Length 196;
Best Local Similarity 100.0%; Pred. No. 7,1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY              3 CGATGATGATTACTT 19
                |||||||
DB              86 CGATGATGATTACTT 102

RESULT 9
AA254497        286 bp      mRNA          linear      EST 14-MAR-1997
DEFINITION      val7r04.r1 Soares mouse lymph node BMDL Mus musculus cDNA clone
IMAGE           723174 5', mRNA sequence.
ACCESSION       AA254497
VERSION         AA254497.1 GI:1889091
KEYWORDS        EST.
SOURCE          Mus musculus (house mouse)
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE       1 (bases 1 to 286)
AUTHORS         Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,U., Morris,M., Seisler,B., Gelboin,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,


```


Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Tissues were provided by Michela Fagiolini and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirosawa, Wako-shi, Saitama 351-0198 Japan) whose assistance we gratefully acknowledge.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES
 source
 location/Qualifiers
 1. 401
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="K230310G10"
 /tissue_type="visual cortex"
 /clone_lib="RIKEN full-length enriched, visual cortex"

ORIGIN

Query Match 81.0%; Score 17; DB 6; Length 401;
 Best Local Similarity 100.0%; Pred. No. 7.9e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CGGATGATGATTACTT 19
 |||||
 Db 304 CGGATGATGATTACTT 320

Search completed: November 7, 2005, 09:18:30
 Job time : 2109.89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:11:16 ; Search time 668.064 Seconds
(without alignments)
1233.024 Million cell updates/sec

Title: US-10-777-131A-2

Perfect score: 17
Sequence: 1 tccgccgggtcgggac 17

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	100.0	17	6	AX770104
2	17	100.0	823	6	AX770121
3	17	100.0	3002	6	CQ880370
4	17	100.0	40056	1	MSGY348
5	17	100.0	110000	1	AE000516_12
6	17	100.0	327650	1	BX248337
7	17	100.0	349306	1	BX842575
8	16	94.1	1566	1	MGU95316
9	16	94.1	134943	2	AC148866
10	16	94.1	300704	2	AE017316
11	16	94.1	314100	1	SC0939106
12	15.4	90.6	356	10	AF133913
13	15.4	90.6	795	3	PVU491803
14	15.4	90.6	1034	10	RATPGP
15	15.4	90.6	1109	8	AK111058
16	15.4	90.6	1194	6	AA8328
17	15.4	90.6	1194	6	AR198358
18	15.4	90.6	1415	6	CQ777515
19	15.4	90.6	1415	10	MMU13705

20	15.4	90.6	1442	10	BC049235	BC049235 Mus muscu
21	15.4	90.6	1448	10	BC003339	BC003339 Mus muscu
22	15.4	90.6	1478	10	BC062227	BC062227 Rattus no
23	15.4	90.6	1495	10	BC061950	BC061950 Mus muscu
24	15.4	90.6	1528	6	BD203690	BD203690 Human nuc
25	15.4	90.6	1528	6	AX014845	AX014845 Sequence
26	15.4	90.6	1652	6	BD203774	BD203774 Human nuc
27	15.4	90.6	1652	6	AX015057	AX015057 Sequence
28	15.4	90.6	1762	10	BC019550	BC019550 Mus muscu
29	15.4	90.6	1775	1	AF224501	AF224501 Streptomy
30	15.4	90.6	2637	4	AB112589	AB112589 Fells cat
31	15.4	90.6	4496	6	AA8326	AA8326 Sequence 4
32	15.4	90.6	4496	6	AR198356	AR198356 Sequence
33	15.4	90.6	5945	1	AY426768	AY426768 Streptomy
34	15.4	90.6	5985	1	SCU43704	SCU43704 Streptomy
35	15.4	90.6	6375	1	AY632768	AY632768 Mycobacte
36	15.4	90.6	8988	1	MTBKATA	MTBKATA
37	15.4	90.6	15065	1	AE005094	AE005094 Halobacte
38	15.4	90.6	70575	2	AC025103	AC025103 Homo sapi
39	15.4	90.6	84985	1	SNA278573	SNA278573 Streptomy
40	15.4	90.6	110000	1	AE016822_01	AE016822_01 of
41	15.4	90.6	110000	2	AE016822_02	AE016822_02 of
42	15.4	90.6	110000	2	AC106698_2	AC106698_2 of
43	15.4	90.6	110000	2	LMFLCHR15_6	LMFLCHR15_6 of
44	15.4	90.6	129768	2	AC134955	AC134955 Tetradon
45	15.4	90.6	138809	2	AC151220	AC151220 Bos tauru

ALIGNMENTS

RESULT 1	AX770104	17 bp	DNA	linear	PAT 02-JUL-2003
LOCUS	AX770104	Sequence 2 from Patent WO03016562.			
DEFINITION	AX770104				
ACCESSION	AX770104.1	GI:32437682			
VERSION	AX770104.1				
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					
ORIGIN					
Query Match					
Best Local Similarity	100.0%;	Score 17;	DB 6;	Length 17;	
Matches	17;	Conservative	0;	Mismatches	0;
Indels				Gaps	0;
Qy	1	TCCGCCGGGTGCGGAC	17		
Db	1	TCCGCCGGGTGCGGAC	17		
RESULT 2					
LOCUS	AX770121	823 bp	DNA	linear	PAT 02-JUL-2003
DEFINITION	Sequence 19 from Patent WO03016562.				
ACCESSION	AX770121				
VERSION	AX770121.1	GI:32437699			
KEYWORDS					
SOURCE					
ORGANISM					

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1
AUTHORS Gicquel, B.
TITLE Compositions and methods for detecting multidrug resistant strains of M. tuberculosis having mutations in genes of the mult family
JOURNAL Patent: WO 03016562-A 19 27-FEB-2003;
INSTITUT PASTEUR (FR)
FEATURES Location/Qualifiers
source 1. 823
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"

ORIGIN
Query Match 100.0%; Score 17; DB 6; Length 823;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCCGCCGGGTCTGGGAC 17
Db 820 TCCGCCGGGTCTGGGAC 804

RESULT 3
LOCUS CO880370 3002 bp DNA linear PAT 11-OCT-2004
DEFINITION Sequence 1 from Patent WO2004083459.
ACCESSION CO880370
VERSION CO880370.1 GI:54034136
KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1
AUTHORS Bange, F.C.
TITLE Method and kit for a specific detection of m. tuberculosis
JOURNAL Patent: WO 2004083459-A 1 30-SEP-2004;
Arctus Gesellschaft Fuer Molekularbiologische Diagnostik und Entwicklung MBH. (DE)
FEATURES Location/Qualifiers
source 1. 3002
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"

ORIGIN
Query Match 100.0%; Score 17; DB 6; Length 3002;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCCGCCGGGTCTGGGAC 17
Db 1603 TCCGCCGGGTCTGGGAC 1587

RESULT 4
MSGY348 40056 bp DNA linear BCT 10-DEC-1996
LOCUS MSGY348
DEFINITION Mycobacterium tuberculosis sequence from clone y348.
ACCESSION AD000020
VERSION AD000020.1 GI:11717739
KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1 (bases 1 to 40056)
Du, L.

TITLE Direct Submission
JOURNAL Submitted (11-OCT-1996) L. Du, Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA, USA, 02154 du@gtc.com
COMMENT GSDB:S:1004716.
FEATURES Location/Qualifiers
source 1. 40056
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"
/clone="y348"

ORIGIN
Query Match 100.0%; Score 17; DB 1; Length 40056;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCCGCCGGGTCTGGGAC 17
Db 18734 TCCGCCGGGTCTGGGAC 18718

RESULT 5
AE000516_12/c
WPCOMMENT
Sequence split into 44 fragments LOCUS AE000516 Accession AE000516
Fragment Name Begin End
AE000516_00 1 110000
AE000516_01 100001 210000
AE000516_02 200001 310000
AE000516_03 300001 410000
AE000516_04 400001 510000
AE000516_05 500001 610000
AE000516_06 600001 710000
AE000516_07 700001 810000
AE000516_08 800001 910000
AE000516_09 900001 1010000
AE000516_10 1000001 1110000
AE000516_11 1100001 1210000
AE000516_12 1200001 1310000
AE000516_13 1300001 1410000
AE000516_14 1400001 1510000
AE000516_15 1500001 1610000
AE000516_16 1600001 1710000
AE000516_17 1700001 1810000
AE000516_18 1800001 1910000
AE000516_19 1900001 2010000
AE000516_20 2000001 2110000
AE000516_21 2100001 2210000
AE000516_22 2200001 2310000
AE000516_23 2300001 2410000
AE000516_24 2400001 2510000
AE000516_25 2500001 2610000
AE000516_26 2600001 2710000
AE000516_27 2700001 2810000
AE000516_28 2800001 2910000
AE000516_29 2900001 3010000
AE000516_30 3000001 3110000
AE000516_31 3100001 3210000
AE000516_32 3200001 3310000
AE000516_33 3300001 3410000
AE000516_34 3400001 3510000
AE000516_35 3500001 3610000
AE000516_36 3600001 3710000
AE000516_37 3700001 3810000
AE000516_38 3800001 3910000
AE000516_39 3900001 4010000
AE000516_40 4000001 4110000
AE000516_41 4100001 4210000
AE000516_42 4200001 4310000
AE000516_43 4300001 4403837
Continuation (13 of 44) of AE000516 from base 1200001 (AE000516 Mycobacterium tuberculosis
Query Match 100.0%; Score 17; DB 1; Length 110000;

Best Local Similarity 100.0%; Pred. No. 9.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCGCCGGGTCCGGGAC 17
|||||
Db 86702 TCCGCCGGGTCCGGGAC 86686

RESULT 6
BX248337/c 327650 bp DNA linear BCT 06-MAY-2004
LOCUS Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment
DEFINITION 4/14.
ACCESSION BX248337 BX248333
VERSION BX248337.1 GI:31617663
KEYWORDS complete genome.
SOURCE Mycobacterium bovis AF2122/97
ORGANISM Mycobacterium bovis AF2122/97
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacteriineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE 1
AUTHORS Garnier, T., Eigleier, K., Camus, J.-C., Medina, N., Mansoor, H.,
Pryor, M., Dutfoy, S., Grondin, S., Lacroix, C., Monsempé, C., Simon, S.,
Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L.,
Hewinson, P. R., Parkhill, J., Barrall, B. G., Cole, S. T., Gordon, S. V. and
The complete genome sequence of Mycobacterium bovis
Unpublished
2 (bases 1 to 327650)
Garnier, T.
Direct Submission

JOURNAL Submitted (24-MAR-2003) Garnier T., Unité de Génétique Moléculaire
Bactérienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex
15, France. e-mail: cgarnier@pasteur.fr Submitted on behalf of the
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
Laboratories Agency Weybridge, Wobham Lane, New Haw, Addlestone,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK. Pr4 Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28
rue du Docteur Roux, 75724 Paris Cedex 15, France
Location/Qualifiers

FEATURES
source 1..327650
/organism="Mycobacterium bovis AF2122/97"
/mol_type="genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"

gene 698..2170
/locus_tag="Mb0912"
698..2170
/locus_tag="Mb0912"
698..2170
/locus_tag="Mb0912"
/note="Mb0912", -, len: 490 aa. Equivalent to Rv0888, len:
490 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 490 aa overlap). Probable exported
protein. Equivalent to AKK45157.1 from Mycobacterium
tuberculosis strain CDC1551 (507 aa) but shorter 17 aa.
Contains possible N-terminal signal sequence."
/codon_start=1
/transl_table=1
/product="PROBABLE EXPORTED PROTEIN"
/protein_id="CAD93773.1"
/db_xref="GI:31617664"
/db_xref="GOA:Q10549"
/db_xref="UniProt/Swiss-Prot:Q10549"
/translation="MDYAKRIGQVGALVAVGVGAATVTHAIGSAAPTPDSSSTDS
VDACPLGGSASSLAIPGASVPGVGVQVDPGSI.PDDLLNALIDFLAVNRGLVPII
ENRTPANPQGVSEGGTGVPRDADCPDGNRMTFAVRERGAQGHQGLVVDOR
TASFIYADPGVGTDSVNVSDTSLHVLGAGLYDPFHGDHVAATVAVNGTPT
DTIGSDPSMLTYNAGLPLPSSALILPRFPYTKELIGKRLNYVAVNADPAHYOFLI
KSKRPSQTPPEPTLILMIGVPSDGLANTISERFQRLDRQTYECTSDNCLTKGF
TYSQRLPGSDGVNLTNTGSGPTTNANLAQVANYIQNSAGRAVIVGDENARY
SDDSALLQFAQVNGLTDAWVQVEHGPPTTPPAFCWNGNECELDKIFYSGGQVTL

gene

CDS

OAVSGNEAPKEFNSKCEPLSDHSPAVVGFHYADNVAVR"
complement(12205..3326)
/gene="cItA"
/locus_tag="Mb0913c"
complement(12205..3326)
/gene="cItA"
/locus_tag="Mb0913c"
/EC_number="2.3.3.1"
/note="Mb0913c, cItA, len: 373 aa. Equivalent to Rv0893c,
len: 373 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 373 aa overlap). Probable cItA
(alternate gene name: gItA), citrate synthase 2 (EC
4.1.3.7), highly similar to others e.g.
CAB95899.1|AL35998 putative citrate synthase from
Streptomyces coelicolor (367 aa); P39191|CISY_BACSU
citrate synthase II from Bacillus subtilis (366 aa), FASTA
scores: opt: 586, E(): 5.8e-30, (33.8% identity in 367 aa
overlap); etc. Also similar to Rv0896|MTCY31.24 from
Mycobacterium tuberculosis (29.2% identity in 274 aa
overlap) and Rv1131. Contains P500480 Citrate synthase
signature. BELONGS TO THE CITRATE SYNTHASE FAMILY."
/codon_start=1
/transl_table=1
/product="PROBABLE CITRATE SYNTHASE II CITA"
/protein_id="CAD93774.1"
/db_xref="GI:31617665"
/db_xref="GOA:Q10529"
/db_xref="UniProt/Swiss-Prot:Q10529"
/translation="MTVPENFVPGLDGVAFETIELEPPDKGALRYGVDIEDLV
QRTVFGVWALVDVNGFSGSLPPAPPEPLPHSGDVVDVQAGLMLAPINGVAPLD
IDDAFARQQLARASVMAISYVAGARGIYQVAPQRIIDECSTIARMTWQSEPD
RHIEADIAQVWASAEGHGNASTFTFARVYASGADVAAALSGAIGMSPGLHGPARY
LPMIDDEVARACDASVYKGLIDREKXMGFGRVRYRAEDPPARVLRRAERLGPARY
VAVAEQALSELSERRPRDRIETNVEFMAVVIDPARVPANMPAMFTCGRTAGWCA
HILEQKRLGKLVRSALYVGGPSPSSVDSEMERVLTAA"
complement(3413..6061)
/locus_tag="Mb0914c"
complement(3413..6061)
/locus_tag="Mb0914c"
/note="Mb0914c", -, len: 882 aa. Equivalent to Rv0890c,
len: 882 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.9% identity in 882 aa overlap). Probable
transcriptional regulatory protein, luxR family, highly
similar (but shorter 238 aa in N-terminus) to
NP_302202.1|NC_002677 possible transcriptional regulator
from Mycobacterium leprae (1106 aa). Also highly similar
(generally in part) to others e.g. T50568 probable
multi-domain regulatory protein from Streptomyces
coelicolor (1334 aa); P10957|NARL_ECOLI nitrate/nitrite
response regulator protein from Escherichia coli (216 aa)
FASTA scores: opt: 193, E(): 6e-06, (37.4% identity in 99
aa overlap); etc. Also highly similar to others from
Mycobacterium tuberculosis e.g. MTGY02B10.22, MTV008.44,
MTV036.21, and MTCY31.24. Contains P500017 ATP/GTP-binding
site motif A (P-loop), P500622 Bacterial regulatory
proteins, luxR family signature, and probable helix-turn
helix motif from aa 836 to 857 (Score 1559, +4.50 SD).
BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
REGULATORS."

gene
CDS

/transl_table=1
/product="PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN
(PROBABLY LUXR-FAMILY)"
/protein_id="CAD93775.1"
/db_xref="GI:31617666"
/db_xref="UniProt/Swiss-Prot:P59699"
/translation="IMRALIQNRILVLTCTGCGVQKRTLAIQIASASELRDGLCFVDLA
PITSGSLVAATAAVGLPDQPGSTDSLR.FGRNMLAVLNCEHLDAACALV
ELDAGCELTILARSIPGAGBITRVSPLSTDEVELFARASVQGFYAH
NAAVGICERLQISQIPAIKFAARVASMPLSTIDDELDFLAGGVRAVQROOT
LRASIDMSHALITEOILRRLRAPFGGPDLAVRVAAGSDIDPPSVLDQULTLV
KSLVADDCQRTRETLFVRLVRLKLDGSGADVAHRRADYTLAASLNPADN
DHQRLVARFETIDNLRAPAFWSRENCHITLQLASLQPIWGRALHRLGSLWFNS

IIEDORFRLAVSTAVPARALADRAMLSTMTATSPVAGATTIIPAQOALAMAREVGP
AALVRLATACGSSGYNAEAPAFPAEATDIAIDMTGTCOILYRGVGTCSGDP
NALBAAREEDLADTTGDRVSRHSCMSLMMNGMTLEALEIRETAEASN
DUPKRVGLYTOAOVUAYCGASAAHATAGTAAATEIGCYOICGVAAMTYALAG
DVTALAEASDARPTILRAQPDQVTHVOYMAQALHAGDDATAAOFADADATNGM
RMVALTTARVATARGEBELADDAHAAAGASLHYOGMDHEMLTAGVSGSH
SEBALVGTAAALAQOTQVRFKIMDAVOYASVALRAMEDEDFDRMAGCALSTD
EALVAVOGRGERGRPARMGSLTPTEEDVVRLVSEGLSNDIKARLFVSPTVQTHL
THVYAKGLASRVOUVEAARSGSPS
complement (6063..6920)
/locus_tag="Mb0915c"
/locus_tag="Mb0915c"
complement (6063..6920)
/locus_tag="Mb0915c"
/locus_tag="Mb0915c"
/note="Mb0915c", len: 285 aa. Equivalent to RV0891C,
len: 285 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.6% identity in 285 aa overlap). Possible
transcriptional regulator, highly similar in N-terminus to
NP_302202.1|NC_002677 possible transcriptional regulator
from Mycobacterium leprae (1106 aa). Also highly similar
to several Mycobacterium tuberculosis putative
transcriptional regulators e.g. Q1102|MTCY02B10.22
PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN (1159 aa),
FASTA scores: opt: 702, E(): 8.3e-40, (50.6% identity in
247 aa overlap); MT036_21; MT008_44; MTCY02B10_23. Also
shows similarity with several adenylate cyclases and
hydrolases from other organisms."

gene
CDS
/transl_table=11
/product="POSSIBLE TRANSCRIPTIONAL REGULATORY PROTEIN"
/protein_id="CAD93776.1"
/db_xref="GI:31617667"
/db_xref="UniProt/Swiss-Prot:P59970"
/translation="MLFNAVNSLPNIDIALIRGEDHPGCAKCYARGISALGS
LDLYHSIRCYAAPDVGRCRFPERRRVLVANGDLSRIPLDPTLTLLADVEST
HLMQCPEDMAFAIAHDHTYSBATINHGQYQPKRYGDSFVAATRASAAACAD
LQRTSLARLRIGHTGSEVQRLDELTYGPTINRARKLDAHGGVLSATGDLVT
GLPADAVLVDLGRHPLRGLPRPWWQDCHPDIREKPPRLTAKSSPTSLPQPTT
FVGRRAQIS"
7318..8805
/locus_tag="Mb0916"
/locus_tag="Mb0916"
/locus_tag="Mb0916"
/EC_number="1.14.-.-"
/note="Mb0916", len: 495 aa. Equivalent to RV0892, len:
495 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 495 aa overlap). Probable
monooxygenase (EC 1.14.-.-), highly similar to others e.g.
NP_250787.1|NC_002516 probable flavin-binding
monooxygenase from Pseudomonas aeruginosa (491 aa);
CAB59668.1|AL132674 monooxygenase from Streptomyces
coelicolor (519 aa); P12015|CYMO_ACIS cyclohexanone
monooxygenase from Actinobacter sp. (542 aa). FASTA
scores: opt: 489, E(): 6.8e-26, (30.3% identity in 492 aa
overlap); etc. Also highly similar to RV0565c, RV3854c,
RV3083, etc from Mycobacterium tuberculosis. Has
hydrophobic stretch at N-terminus."
/codon_start=1
/transl_table=11
/product="PROBABLE MONOOXYGENASE"
/protein_id="CAD93777.1"
/db_xref="GI:31617668"
/db_xref="GOA:Q10532"
/db_xref="UniProt/Swiss-Prot:Q10532"
/translation="MTGRCPYAVVAGAMSGCAVITLLSAGITDVCIYERADVGT

Query Match 100.0%; Score 17; DB 1; Length 327650;
Best Local Similarity 100.0%; Pred. No. 7.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCCGCGGGTCCGGGAC 17
|||||
DB 301586 TCCCGCGGGTCCGGGAC 301570

RESULT 7
BX842575/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE

1
Cole S.T., Brosh R., Parkhill J., Garnier T., Churcher C.,
Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry III C.E.,
Tekaia F., Badcock K., Basham D., Brown D., Chillingworth T.,
Comnor R., Davies R., Devlin K., Feltwell T., Gentles S.,
Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogan A., McLean J.,
Moule S., Murphy L., Oliver S., Osborne J., Quail M.A.,
Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S.,
Squares S., Squires R., Sulston J.E., Taylor K., Whitehead S. and
Barrell B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (6685), 537-544 (1998)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/).

FEATURES

SOURCE

gene
CDS

1. 349306
/organism="Mycobacterium tuberculosis H37Rv"
/mol_type="genomic DNA"
/strain="H37Rv"
/db_xref="taxon:83332"
complement (53..844)
/locus_tag="RV0927c"
complement (53..844)
/locus_tag="RV0927c"
/EC_number="1.-.-.-"
/function="UNKNOWN; POSSIBLY INVOLVED IN CELLULAR
METABOLISM."
/note="RV0927c, (MTCY21C12.21c), len: 263 aa. Probable
short-chain dehydrogenase/reductase (EC 1.-.-.-), similar
to various dehydrogenases/reductases, notably
7-alpha-hydroxysteroid dehydrogenase and glucose
1-dehydrogenases e.g. P25529|HHA_ECOLI
7-alpha-hydroxysteroid dehydrogenase from Escherichia coli
(255 aa), FASTA scores: opt: 551, E(): 1e-26, (19.5%
identity in 248 aa overlap); NP_252778.1|NC_002516

probable short-chain dehydrogenase from *Pseudomonas aeruginosa* (253 aa); AAC44307.1|US9433.3-*ketacyl-acyl* carrier protein reductase from *Bacillus subtilis* (246 aa); etc. Also similar to other dehydrogenases from *Mycobacterium tuberculosis* e.g. MTCY09F9.36, E():1.4e-18; MTCY369.14, E():8e-17; MTCY02B10.14, E():2.5e-14; MTCY09F9.23c, E():1.5e-13; MTCY03C7.07, E():1.9e-13. Contains PS00061 Short-chain dehydrogenases/reductases family signature, and PS00017 ATP/GTP-binding site motif A (P-loop). BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY."

/transl_table=1
/codon_start=1
/product="PROBABLE SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE"
/protein_id="CAB08497.1"
/db_xref="GI:2078076"
/translation="MILDMFLDDKAVITGGGNGAIALAPQAQADVLASRTS SELDAVQIRAGRAHVAADLAPVLAAGVAFKGLDIVNNVNGTMENT LLSTKDLADFAFNVTGAHLTVAAVPLMEHSGGSGVINISITMRAGRPAAV GTAKALAHYRIALDI.CPRVRVNAIAPGSLTSLALVVAANDELRAPEMCAITLR LGDPVIAAAVYIAPSLPAGSFLTKTLEVDGGLTFPNDLPIIDL"

/complement(113..136)
/locus_tag="Rv0927c"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
/complement(320..406)
/locus_tag="Rv0927c"
/note="PS00061 Short-chain alcohol dehydrogenase family signature"
1116..2228
/gene="pbcS3"
/locus_tag="Rv0928"
1116..2228
/gene="pbcS3"
/locus_tag="Rv0928"
/function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT). THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE TRANSPORT."

/standard_name="phoS2"
/note="RV0928, (MTCY21C12.22), len: 370 aa. pbcS3 (previously known as phoS2), phosphate-binding lipoprotein component of inorganic phosphate transport system (see citations below), highly similar to others from *Mycobacterium leprae* e.g. Q50099|PBCS3|PHOS1 phosphate-binding protein 3 precursor (328 aa), PASTA scores: opt: 1772, E(): 0, (79.6% identity in 328 aa overlap); and highly similar to others e.g. AAF74819.1|AF137360.1|AF137360 periplasmic phosphate permease from *Mycobacterium avium* (369 aa). Also highly similar to Rv0932c|MTCY08D9.07|pbcS2 PHOSPHATE-BINDING PERIPLASMIC LIPOPROTEIN (370 aa); and Rv0934|pbcS1 PHOSPHATE-BINDING PERIPLASMIC LIPOPROTEIN (374 aa) from *Mycobacterium tuberculosis* (*Mycobacterium tuberculosis* seems to have three pbcS-like proteins, others being Rv0932c and Rv0934c). Contains lipoprotein signature (PS00013) at N-terminus. BELONGS TO FAMILY OF PHOSPHATE RECEPTORS FOR BACTERIAL ABC-TYPE LIPOPROTEIN TRANSPORTERS."

/transl_table=1
/evidence=experimental
/codon_start=1
/product="PERIPLASMIC PHOSPHATE-BINDING LIPOPROTEIN PBCS3 (PBP-3) (PBCS3) (PHOS1)"
/protein_id="CAE5337.1"
/db_xref="GI:41353620"
/translation="MKLNRFQAAVGVLAAGALVLSACGNDNVYTGATVGAQAKV CGGKTLKASGTAQANMTRPVNFEQACGGOTLNTYANGSAGISRFNGNDPFG SDVPLSKDEAAAOGRGSPAMNLPLVPGFALVATYVNLNSVLSLIDGTLAKIKNGSI TOMNPAIQALNDRFTLBERGRIHYVFRDESGTIDNFRQYIQAASNGWAGAGASFC GVGEGANGDGTSLAAKNTPGSLITYNWSFRAQDLHLMANVTSAGGDPAITIDSV GQTLAGATISGVDGLVLDTSFVRPKRPGSYPIVLAIVETVCSKYPSDVGTAVKAR LQSTIGAGQSGIDNGYIPIIDPEFKSRSLTAVNMA"

misc_feature 1152..1184
/gene="pbcS3"
/locus_tag="Rv0928"
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
2241..3215
/gene="pbcS2"
/locus_tag="Rv0929"
2241..3215
/gene="pbcS2"
/locus_tag="Rv0929"
/function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE. THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE TRANSPORT."
/note="Rv0929, (MTCY21C12.23), len: 324 aa. pbcS2, phosphate-transport integral membrane ABC transporter (see citations below), highly similar to others e.g. NP_302394.1|NC_002677 membrane-bound component of phosphate transport from *Mycobacterium leprae* (319 aa); CAB88474.1|ALJ53816 phosphate ABC transporter system permease protein from *Streptomyces coelicolor* (336 aa); NP_290359.1|NC_002655 high-affinity phosphate-specific transport system (cytoplasmic membrane component) from *Escherichia coli* strain O157:H7 (319 aa); etc. Also similar to Rv935|MTCY08D9.04c|PBC1 PROBABLE TRANSMEMBRANE ABC TRANSPORTER COMPONENT OF PHOSPHATE UPTAKE SYSTEM from *Mycobacterium tuberculosis* (338 aa). Contains binding-protein-dependent transport systems inner membrane component signature (PS00402)."
/codon_start=1
/transl_table=1
/product="PHOSPHATE-TRANSPORT INTEGRAL MEMBRANE ABC TRANSPORTER PBC2"
/protein_id="CAB08489.1"
/db_xref="GI:3261752"
/translation="MTVEPLTKPALVAVDMRPARGRFLFLAASGTTIVAILI AIFLVAVPSLRNANHFSTPTSDSQALFGRDLEFMTALSGITLVLAVP AVGIAVFLTHYAPRLSRPGAWDLAAVPSII.FGLMGIFVLAPKLEIPAFRLNL GMLPLFKQGVNSLAGGTTIPAGIVLSVLMPLTYSREVFRTPLQIOIAALALGA TKMVLVMTVLPYGRSGVVAASMGIRALGDETVAVLVILASARPWTSLFDGGYTF ASKIASAASESEPLPGAYISAGFALVFLFLVNAABAIAAGKVNQ"
2844..2930
/gene="pbcS2"
/locus_tag="Rv0929"
/note="PS00402 Binding-protein-dependent transport systems inner membrane comp signature"
3212..4138
/gene="pbcA1"
/locus_tag="Rv0930"
3212..4138
/gene="pbcA1"
/locus_tag="Rv0930"
/function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE."

Query Match 100.0%; Score 17; DB 1; Length 349306;
Best Local Similarity 100.0%; Pred. No. 7.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCGCCGGGTGGGGAC 17
|||||
Db 253427 TCCGCCGGGTGGGGAC 253411
|||||

RESULT 8
LOCUS MGU95316 1566 bp DNA linear BCT 11-MAY-1997
DEFINITION *Mycobacterium gordonae* IS1512 transposase gene, complete cde, and Tn554 tnpB transposase homolog gene, partial cde.
ACCESSION U95316

```

VERSION      U95316.1  GI:2078346
KEYWORDS
SOURCE       Mycobacterium gordonae
ORGANISM     Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium.
REFERENCE    1 (bases 1 to 1566)
              Picardeau, M. and Vincent, V.
              Identification and characterization of IS-like elements in
              Mycobacterium gordonae
              Unpublished
              2 (bases 1 to 1566)
              Picardeau, M. and Vincent, V.
              Direct Submission
              Submitted (27-MAR-1997) Laboratoire de Reference des Mycobacteries,
              Institut Pasteur, 25 rue du Docteur Roux, Paris 75013, France
FEATURES
  source
    1..1566
      /organism="Mycobacterium gordonae"
      /mol_type="genomic DNA"
      /strain="960592"
      /db_xref="taxon:1778"
    1..>893
      /organism="Mycobacterium gordonae"
      /mol_type="genomic DNA"
      /strain="960592"
      /db_xref="taxon:1778"
      /insertion_seq="IS1512"
      /note="IS256 family member; truncated"
    54..893
      /codon_start=1
      /transl_table=1
      /product="IS1512 transposase"
      /protein_id="AAB54014.1"
      /db_xref="GI:2078347"
      /translation="MTAAHDIDLPALLAERLTTPDVALRELLATFIHTTNGAERADL
      CGAAYGGRSSERTSRNGYRHOEPDTRAGSLDLAI PKLRQSYFPDMLERKKAERA
      LTTVAATCYLGVSTRMDKLVETGIGTSGQSVVAKELDTAVEAFRRPDAEP
      YTPMAADALVKKREAGRVVVALIANGVADGRELIGIDTTAAGAGMLTFELS
      LTARGSGVALVTSDAHAGLLSALGATLPGASMKRCRTHVATINRPSLBSRRQVGRG
      VAA"
      complement(858..>1566)
      /note="similar to tm554 tpm transposase"
      /codon_start=2
      /transl_table=1
      /product="transposase homolog"
      /protein_id="AAB54015.1"
      /db_xref="GI:2078348"
      /translation="WITGRGRQRLPHPRGRLTEFLFVEHGCRPTSRRLKGLIDAV
      AAARLGRDGLIHTVPHQLRHTTGTSLINGIGI.PALMALMGHTVTPMTLRKLAT
      PTIRSAVQVAMDKVAGQLPLTAVNAAPIPDKVAVMIAEMLKTRLAHGFCARPOA
      GPCPYANI CECDCDFVPDPAGTSTINQINIRALQADQAQRGMHDEAARQRVAASL
      DQHLQRLRDRSTDLPS"
  ORIGIN
    Query Match      94.1%; Score 16; DB 1; Length 1566;
    Best Local Similarity 100.0%; Pred. No. 6e+03;
    Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    Oy 2 CCGCCGGGTCCGGGAC 17
    Db 1025 CCGCCGGGTCCGGGAC 1040
  RESULT 9
  AC148866 134943 bp DNA linear HTG 08-MAY-2004
  LOCUS     Felis catus clone RP86.360D22, WORKING DRAFT SEQUENCE, 8 ordered
  DEFINITION
  AC148866
  AC148866.2 GI:47084442
  VERSION   HTG; HTGS_PHASE2; HTGS_DRAFT.
  KEYWORDS  Felis catus (cat)
  SOURCE

```

```

ORGANISM     Felis catus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
REFERENCE    1 (bases 1 to 134943)
              Antonellis, A., Ayele, K., Benjamin, B., Blakeley, R. W.,
              Bouffard, G. G., Brinkley, C., Brooks, S., Chu, G., Coleman, B.,
              Coleman, H., Daki, N., Engle, J., Grante, S., Guan, X., Gupta, J.,
              Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Hu, P., Hurle, B.,
              Idol, J. R., Jones, C., Karlins, E., Kim, H., Kwong, P., Latic, P.,
              Larson, S., Lee-Lin, S.-O., Legaspi, R., Maduro, O. L., Maduro, V. B.,
              Margulies, E. H., Masiello, C., Maubert, B., McDowell, J.,
              Mullikin, J. C., Paguirigan, C., Potvin, M. E., Prasad, A., Pur, O.,
              Reddi-Dugue, N., Schandler, K., Schueler, M. G., Shah, K., Sison, C.,
              Stentrop, S., Thomas, J. W., Thomas, P. J., Tsipouri, V., Vogt, J. L.,
              Wehrhapp, K. D., Young, A. and Green, E. D.
              NISC Comparative Sequencing Initiative
              Unpublished
              2 (bases 1 to 134943)
              Green, E. D.
              Direct Submission
              Submitted (14-APR-2004) NIH Intramural Sequencing Center, 8717
              Grovemont Circle, Gaithersburg, MD 20877, USA
              3 (bases 1 to 134943)
              Green, E. D.
              Direct Submission
              Submitted (08-MAY-2004) NIH Intramural Sequencing Center, 8717
              Grovemont Circle, Gaithersburg, MD 20877, USA
              On May 8, 2004 this sequence version replaced gi:46367828.
              ----- Genome Center
              Center: NIH Intramural Sequencing Center
              Center code: NISC
              Web site: http://www.nisc.nih.gov
              Contact: nisc.zoo@nih.gov
              ----- Project Information
              Center project name: efn
              Center clone name: 360D22

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies; the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

```

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 133378 bases at least Q40
Consensus quality: 13780 bases at least Q30
Consensus quality: 134067 bases at least Q20
Insert size: 123000; agarose-fp
Insert size: 134243; sum-of-contigs
Quality coverage: 11.09x in Q20 bases; agarose-fp
Quality coverage: 10.17x in Q20 bases; sum-of-contigs

```

----- NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

```

1 2073 2172: contig of 2072 bp in length
2073 2172: gap of unknown length
2173 6288: contig of 4116 bp in length

```

* 6289 6388: gap of unknown length
* 6389 44101: contig of 37713 bp in length
* 44102 44201: gap of unknown length
* 44202 66179: contig of 21978 bp in length
* 66180 66279: gap of unknown length
* 66280 68096: contig of 1817 bp in length
* 68097 68196: gap of unknown length
* 68197 72556: contig of 4360 bp in length
* 72557 72657: gap of unknown length
* 72657 113988: contig of 41332 bp in length
* 113989 114088: gap of unknown length
* 114089 134943: contig of 20855 bp in length.
Location/Qualifiers
source
1. .134943
/organism="Felis catus"
/mol_type="genomic DNA"
/db_xref="taxon:9685"
/clone="RP86-360D22"
/clone_11b="RP86"
/note="BAC resource: <http://bacpac.chori.org/>"
misc_feature
1. .70717
/note="clone overlaps with GenBank Accession Number AC092518 clone RP86-160J10 (center project name cfl)"
misc_feature
1. .2072
/note="assembly_fragment
clone_end:SP6
vector_side:left"
misc_feature
2173. .6288
/note="assembly_fragment"
misc_feature
6389. .44101
/note="assembly_fragment"
misc_feature
44202. .66179
/note="assembly_fragment"
misc_feature
66280. .68096
/note="assembly_fragment"
misc_feature
68197. .72556
/note="assembly_fragment"
misc_feature
72657. .113988
/note="assembly_fragment"
misc_feature
107778. .134943
/note="clone overlaps with GenBank Accession Number AC091696 clone RP86-117J4 (center project name awg)"
misc_feature
114089. .134943
/note="assembly_fragment
clone_end:T7
vector_side:right"
ORIGIN
Query Match 94.1%; Score 16; DB 2; Length 134943;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 CCGCGGGTGGGAC 17
|||||
Db 5955 CCGCGGGTGGGAC 5940
|||||
RESULT 10
AE017316 300704 bp DNA linear BCT 27-APR-2004
LOCUS AE017316
DEFINITION Desulfotribrio vulgaris subsp. vulgaris str. Hildenborough, section 8 of 12 of the complete genome.
ACCESSION AE017316 AE017285
VERSION AE017316.1 GI:46449852
KEYWORDS
SOURCE
ORGANISM Desulfotribrio vulgaris subsp. vulgaris str. Hildenborough
Bacteria; Proteobacteria; Deltaproteobacteria; Desulfotribriales;
Desulfotribriaceae; Desulfotribrio.
1 (bases 1 to 300704)
Heidelberg, J.F., Seehadri, R., Haveman, S.A., Hemme, C.L.,
Paulsen, I.T., Kolonay, J.F., Eisen, J.A., Ward, N., Methe, B.,
Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Dodson, R.J., Durkin, A.S.,

Madupu, R., Nelson, W.C., Sullivan, S.A., Four, D.E., Haft, D.H.,
Selengut, J., Peterson, J.D., Daviden, T.M., Zafar, N., Zhou, L.,
Radune, D., Dimitrov, G., Hance, M., Tian, K., Khouri, H.M., Gill, J.,
Uteerback, T.R., Feldblyum, T.V., Wall, J.D., Voordouw, G. and
Fraser, C.M.
The genome sequence of the anaerobic, sulfate-reducing bacterium
Desulfotribrio vulgaris Hildenborough
Nat. Biotechnol. (2004) In press
2 (bases 1 to 300704)
Heidelberg, J.F., Seehadri, R., Haveman, S.A., Hemme, C.L.,
Paulsen, I.T., Kolonay, J.F., Eisen, J.A., Ward, N., Methe, B.,
Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Dodson, R.J., Durkin, A.S.,
Madupu, R., Nelson, W.C., Sullivan, S.A., Four, D.E., Haft, D.H.,
Selengut, J., Peterson, J.D., Daviden, T.M., Zafar, N., Zhou, L.,
Radune, D., Dimitrov, G., Hance, M., Tian, K., Khouri, H.M., Gill, J.,
Uteerback, T.R., Feldblyum, T.V., Wall, J.D., Voordouw, G. and
Fraser, C.M.
Direct Submission
Submitted (18-MAR-2004) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
source
1. .300704
/organism="Desulfotribrio vulgaris subsp. vulgaris str. Hildenborough"
/mol_type="genomic DNA"
/strains="Hildenborough"
/sub_species="vulgaris"
/db_xref="taxon:882"
122. .436
/locus_tag="DVU2027"
122. .436
/locus_tag="DVU2027"
/note="identified by Glimmer2; putative"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS96502.1"
/db_xref="GI:46449853"
122. .436
/locus_tag="DVU2027"
122. .436
/locus_tag="DVU2027"
/note="identified by match to protein family HMM PF01527"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAS96503.1"
/db_xref="GI:46449854"
/translation="MGATPAPYRIRRRSTEQSLTAQEDGVILELAARSBAELTSW
LSFGTHALISPGQLRQIQIVQELKAKELVR"
926. .1324
/locus_tag="DVU2029"
926. .1324
/locus_tag="DVU2029"
/note="identified by match to protein family HMM PF01527"
/codon_start=1
/transl_table=11
/product="ISDv2, transposase OrfA"
/protein_id="AAS96504.1"
/db_xref="GI:46449855"
/translation="MSRQSAKEISTVEVTVWQRRRTIAKLRVVESSLPQSVSF
VARKYGAIPNVFPRKMSDGGVAIOADRRVVSVAEKALKKRIIRLRLGKRTM
EVLTKAIDARKKLISRSPLPFEDGSL"
1321. .2163
/locus_tag="DVU2030"
1321. .2163
/locus_tag="DVU2030"
/note="identified by match to protein family HMM PF00665"

CDS	/locus_tag="DVU2035"
	/complement(5870..6157)
	/locus_tag="DVU2035"
	/note="Identified by match to protein family HMM PF05016"
	/codon_start=1
	/transl_table=11
	/product="plasmid stabilization system family protein"
	/protein_id="AA96510.1"
	/db_xref="GI:46449861"
	/translation="MSNRKRWRLALRDMDALMHNKADPAFAHVASRILOACTT LSHSRRGRAGRVPGTRBELVVALPYIVAIRVRDEVHLRTHHRKRS"
	/complement(6147..6386)
	/locus_tag="DVU2036"
	/complement(6147..6386)
	/locus_tag="DVU2036"
	/note="Identified by match to protein family HMM PF01402"
	/codon_start=1
	/transl_table=11
	/product="helix-turn-helix protein, CopG family"
	/protein_id="AA96511.1"
	/db_xref="GI:46449862"
	/translation="MSAVTRLDPENQOLEALKAKATDRTSRWLIADAVRYVEBSGS QVAIAIGLRQADGDPASDBEYVASSPARKGVVE"
	/complement(6465..7439)
	/locus_tag="DVU2037"
	/complement(6465..7439)
	/locus_tag="DVU2037"
	/note="Identified by similarity to SP:P29933"
	/codon_start=1
	/transl_table=11
	/product="GDS protein, putative"
	/protein_id="AA96512.1"
	/db_xref="GI:46449863"
	/translation="MSBHVHTTIIDLDGAVFSGTPSGRIIKGFDPKCAHTPNIDPDVV FHCSDIIVYMDDSPLYVPPTGSGKTSAIKQLARANYPVVDYTGARGLEFPDL VGSLSTRDSMEFYOGPIALAMHGILINLEIDLPATGNGILDSEPLCPENKRA GGITITPHPRFRPAATANTNGSDDNGLCYGTIRONLAFDRFLCMYPAPAEKKA LERRAPMLPPDIRTVNDYANERRLRFMSISGTTGHSLSVTFSTTLRMADLTIVR FOPLAROGIOPALTVALDRAFGFKASKOTRTTLHEAQRLLPFPOHAATGE"
	/complement(7493..7789)
	/locus_tag="DVU2038"
	/complement(7493..7789)
	/locus_tag="DVU2038"
	/note="Identified by match to protein family HMM PF05016"
	/codon_start=1
	/transl_table=11
	/product="plasmid stabilization system family protein"
	/protein_id="AA96510.1"
	/db_xref="GI:46449861"
	/translation="MSNRKRWRLALRDMDALMHNKADPAFAHVASRILOACTT LSHSRRGRAGRVPGTRBELVVALPYIVAIRVRDEVHLRTHHRKRS"
	/complement(6147..6386)
	/locus_tag="DVU2036"
	/complement(6147..6386)
	/locus_tag="DVU2036"
	/note="Identified by match to protein family HMM PF01402"
	/codon_start=1
	/transl_table=11
	/product="helix-turn-helix protein, CopG family"
	/protein_id="AA96511.1"
	/db_xref="GI:46449862"
	/translation="MSAVTRLDPENQOLEALKAKATDRTSRWLIADAVRYVEBSGS QVAIAIGLRQADGDPASDBEYVASSPARKGVVE"
	/complement(6465..7439)
	/locus_tag="DVU2037"
	/complement(6465..7439)
	/locus_tag="DVU2037"
	/note="Identified by similarity to SP:P29933"
	/codon_start=1
	/transl_table=11
	/product="GDS protein, putative"
	/protein_id="AA96512.1"
	/db_xref="GI:46449863"
	/translation="MSBHVHTTIIDLDGAVFSGTPSGRIIKGFDPKCAHTPNIDPDVV FHCSDIIVYMDDSPLYVPPTGSGKTSAIKQLARANYPVVDYTGARGLEFPDL VGSLSTRDSMEFYOGPIALAMHGILINLEIDLPATGNGILDSEPLCPENKRA GGITITPHPRFRPAATANTNGSDDNGLCYGTIRONLAFDRFLCMYPAPAEKKA LERRAPMLPPDIRTVNDYANERRLRFMSISGTTGHSLSVTFSTTLRMADLTIVR FOPLAROGIOPALTVALDRAFGFKASKOTRTTLHEAQRLLPFPOHAATGE"
	/complement(7493..7789)
	/locus_tag="DVU2038"
	/complement(7493..7789)
	/locus_tag="DVU2038"
	/note="Identified by match to protein family HMM PF05016"
	/codon_start=1
	/transl_table=11
	/product="plasmid stabilization system family protein"
	/protein_id="AA96510.1"
	/db_xref="GI:46449861"
	/translation="MSNRKRWRLALRDMDALMHNKADPAFAHVASRILOACTT LSHSRRGRAGRVPGTRBELVVALPYIVAIRVRDEVHLRTHHRKRS"
	/complement(6147..6386)
	/locus_tag="DVU2036"
	/complement(6147..6386)
	/locus_tag="DVU2036"
	/note="Identified by match to protein family HMM PF01402"
	/codon_start=1
	/transl_table=11
	/product="helix-turn-helix protein, CopG family"
	/protein_id="AA96511.1"
	/db_xref="GI:46449862"
	/translation="MSAVTRLDPENQOLEALKAKATDRTSRWLIADAVRYVEBSGS QVAIAIGLRQADGDPASDBEYVASSPARKGVVE"
	/complement(6465..7439)
	/locus_tag="DVU2037"
	/complement(6465..7439)
	/locus_tag="DVU2037"
	/note="Identified by similarity to SP:P29933"
	/codon_start=1
	/transl_table=11
	/product="GDS protein, putative"
	/protein_id="AA96512.1"
	/db_xref="GI:46449863"
	/translation="MSBHVHTTIIDLDGAVFSGTPSGRIIKGFDPKCAHTPNIDPDVV FHCSDIIVYMDDSPLYVPPTGSGKTSAIKQLARANYPVVDYTGARGLEFPDL VGSLSTRDSMEFYOGPIALAMHGILINLEIDLPATGNGILDSEPLCPENKRA GGITITPHPRFRPAATANTNGSDDNGLCYGTIRONLAFDRFLCMYPAPAEKKA LERRAPMLPPDIRTVNDYANERRLRFMSISGTTGHSLSVTFSTTLRMADLTIVR FOPLAROGIOPALTVALDRAFGFKASKOTRTTLHEAQRLLPFPOHAATGE"
	/complement(7493..7789)
	/locus_tag="DVU2038"
	/complement(7493..7789)
	/locus_tag="DVU2038"
	/note="Identified by match to protein family HMM PF05016"
	/codon_start=1
	/transl_table=11
	/product="plasmid stabilization system family protein"
	/protein_id="AA96510.1"
	/db_xref="GI:46449861"
	/translation="MSNRKRWRLALRDMDALMHNKADPAFAHVASRILOACTT LSHSRRGRAGRVPGTRBELVVALPYIVAIRVRDEVHLRTHHRKRS"
	/complement(6147..6386)
	/locus_tag="DVU2036"
	/complement(6147..6386)
	/locus_tag="DVU2036"
	/note="Identified by match to protein family HMM PF01402"
	/codon_start=1
	/transl_table=11
	/product="helix-turn-helix protein, CopG family"
	/protein_id="AA96511.1"
	/db_xref="GI:46449862"
	/translation="MSAVTRLDPENQOLEALKAKATDRTSRWLIADAVRYVEBSGS QVAIAIGLRQADGDPASDBEYVASSPARKGVVE"
	/complement(6465..7439)
	/locus_tag="DVU2037"
	/complement(6465..7439)
	/locus_tag="DVU2037"
	/note="Identified by similarity to SP:P29933"
	/codon_start=1
	/transl_table=11
	/product="GDS protein, putative"
	/protein_id="AA96512.1"
	/db_xref="GI:46449863"

TITLE	Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)
JOURNAL	Nature 417 (6885), 141-147 (2002)
MEDLINE	21996410
REFERENCE	12000953
AUTHORS	2 (bases 1 to 314100)
JOURNAL	Bentley, S.D.
TITLE	Direct Submission
COMMENT	Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk On or before Oct 26, 2002 this sequence version replaced gi:20520920, gi:20520898, gi:20520901, gi:20520903, gi:20520904, gi:20520905, gi:20520906, gi:20520907, gi:20520926, gi:20520909, gi:20520910, gi:20520805, gi:20520706, gi:20520789, gi:20520843.
FEATURES	Location/Qualifiers
source	1. 314100
	/organism="Streptomyces coelicolor A3(2)"
	/mol_type="genomic DNA"
	/strain="A3(2)"
	/db_xref="taxon:100226"
RBS	53..56
	67..945
	/gene="SCO0550"
	/note="synonym: SCF11.30"
CDS	67..945
	/gene="SCO0550"
	/note="synonym: SCF11.30"
	67..945
	/gene="SCO0550"
	/note="SCF11.30, possible membrane protein, len: 292 aa; similar to TR:Q9ZBQ4 (EMBL:AL034447) Streptomyces coelicolor putative transmembrane protein, 396 aa; fasta scores: opt: 677 z-score: 505.6 E(): 8.6e-21; 43.7% identity in 309 aa overlap. Contains possible hydrophobic membrane spanning region"
	/codon_start=1
	/transl_table=11
	/product="putative membrane protein"
	/protein_id="CAB59607.1"
	/db_xref="GI:6137052"
	/db_xref="GOA:Q9RK60"
	/db_xref="UniProt/TREMBL:Q9RK60"
	/translation="MGSRRARPSSRHAPRGSLRSRGAALGAVTLTAVGTLTAVGDDGASDRYTNANAADAASNAEBAAGAAFLAASASPSPTAGRSSPSAEKRRPTKTSAPRTSAPKASAPKAPVATKSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSTGSAEAOVATLVNERAAGSPVTANSRLTRADDSDVMASSGVNHSRGPDSSTASVREAGYQMSALGENIAQGGADASVMEAMNSPGRHENTLNCSPYELGVGHFCGCGPMTQDFTGTR"
gene	complement(1031..2359)
	/gene="SCO0551"
	/note="synonym: SCF11.31c"
CDS	complement(1031..2359)
	/gene="SCO0551"
	/note="SCF11.31c, possible histidine kinase protein, len: 442 aa; similar to TR:O50497 (EMBL:AL020958) Streptomyces coelicolor two-component system sensor, 481 aa; fasta scores: opt: 401 z-score: 451.3 E(): 9e-18; 30.2% identity in 367 aa overlap. Contains Pfam matches to entries PF00512 signal, histidine kinase and PF00672 DUF5, domain found in bacterial signal proteins. Contains possible N-terminal signal peptide sequence"
	/codon_start=1
	/transl_table=11
	/product="putative histidine kinase protein"
	/protein_id="CAB59608.1"
	/db_xref="GI:6137053"
	/db_xref="GOA:Q9RK59"
	/db_xref="UniProt/TREMBL:Q9RK59"
	/translation="MRRLAGIALAVTSMVALSPFLVALIVREQARDVTTAAEORAALSPVALITRRADVOQAAVEAGSADOLVRLPDGGVGFPHABEELDAVAGRETLVADTDGMAVLOPVLRGDRVAVEAVPADLTRGVTAAGWALLAIVAGSVLVADRTAPVVRSSRGIKRASIALGSGDLCVRYDPGPELQAGAAPYMAADRYDILAYREVLADISRLRTPTLVLTLEADRMKATPSALRTVBAQGLEREIDSTIAARTPLAAQVLAAGTACPCDVAEVAARKDLFWSVLAQQDRPFERSLTPRPAAVPSDGLAAVVALGNVFRHTPGTRFVAVVERDRHRLTVLDAGGVADPRPALTRAVSDG
misc_feature	complement(1070..1672)
	/gene="SCO0551"
	/note="Pfam match to entry PF00512 signal, Histidine kinase, score 39.60, E-value 7.1e-08"
misc_feature	complement(1712..1921)
	/gene="SCO0551"
	/note="Pfam match to entry PF00672 DUF5, Domain found in bacterial signal proteins, score 36.90, E-value 4.5e-07"
gene	complement(2356..3072)
	/gene="SCO0552"
	/note="synonym: SCF11.32c"
CDS	complement(2356..3072)
	/gene="SCO0552"
	/note="SCF11.32c, possible response regulator, len: 238 aa; similar to TR:CAB54568 (EMBL:MD006392) Streptococcus pneumoniae response regulator Rro2, 234 aa; fasta scores: opt: 583 z-score: 693.9 E(): 2.8e-31; 43.2% identity in 227 aa overlap. Contains Pfam matches to entries PF00486 trans_reg_C, transcriptional regulatory protein and PF00072 response_reg, response regulator receiver domain"
	/codon_start=1
	/transl_table=11
	/product="putative response regulator"
	/protein_id="CAB59609.1"
	/db_xref="GI:6137054"
	/db_xref="GOA:Q9RK58"
	/db_xref="UniProt/TREMBL:Q9RK58"
	/translation="WTDRLTCLPMSATLVVEDDVYTAALILEVTGCHYAVKTHQGFELKRLDTQSPSPVIVLDLGLPDLGDLVLRIRKISRPVPLVATARDDEITRLNLAGADDDVVRPESGGQAAARLAALVLRSAPEGRAPLRVADRIDPTARTHLARLEPLRREFDLAYLAHADQVMSRORLAEVWQQPYVEDQTVVHLSALRKGEXARLRPLRHTVRGIGIKLVAP"
misc_feature	complement(2380..2595)
	/gene="SCO0552"
	/note="Pfam match to entry PF00486 trans_reg_C, transcriptional regulatory protein, C terminal, score 93.00, E-value 1.7e-26"
misc_feature	complement(2710..3042)
	/gene="SCO0552"
	/note="Pfam match to entry PF00072 response_reg, Response regulator receiver domain, score 123.00, E-value 5.6e-33"
RBS	complement(3090..3093)
	complement(3099..3536)
gene	/gene="SCO0553"
	/note="synonym: SCF11.33c"
CDS	complement(3099..3536)
	/gene="SCO0553"
	/note="SCF11.33c, unknown, len: 145 aa"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein SCF11.33c"
	/protein_id="CAB59610.1"
	/db_xref="GI:6137055"
	/db_xref="UniProt/TREMBL:Q9RK57"
	/translation="WKPMFGTABIRHTPSPRENTVALPPSEBGTYTPAPSSHPGAPAPAGSGAVVRGLGVGVTVASGSLAVNDGGDLVRLVGBELGAVTLDTVGKGGAGSSFGAVCATGAVGRVDPPTKWIIVMTAVTLAAVDSQIMR"
gene	complement(3691..5388)
	/gene="SCO0554"
	/note="synonym: mapA, SCF11.34c"
CDS	complement(3691..5388)
	/gene="SCO0554"
	/note="SCF11.34c, manA, probable secreted beta-mannosidase, len: 327 aa; similar to many e.g. MANA_STRLI P51529 mannan endo-1,4-beta-mannosidase precursor (EC 3.2.1.78) (363 aa), fasta scores: opt: 1189 z-score: 1332.7 E(): 0.611% identity in 288 aa overlap. Contains N-terminal signal sequence, and Pfam match to entry PF00150 cellulase, Cellulase (glycosyl) hydrolase family 5). Also similar to SC2H4.16 (62.2% identity in 288 aa overlap)
	/codon_start=1
	/transl_table=11
	/product="putative cellulase, partial CDS, len: 327 aa; similar to many e.g. MANA_STRLI P51529 mannan endo-1,4-beta-mannosidase precursor (EC 3.2.1.78) (363 aa), fasta scores: opt: 1189 z-score: 1332.7 E(): 0.611% identity in 288 aa overlap. Contains N-terminal signal sequence, and Pfam match to entry PF00150 cellulase, Cellulase (glycosyl) hydrolase family 5). Also similar to SC2H4.16 (62.2% identity in 288 aa overlap)

>278 aa; C-terminus similar to SM-GUN4 THEFU (EMBL:U20093)
Thermomonospora fusca endoglucanase E-4 precursor (EC
3.2.1.4) (endo-1,4-beta-glucanase E-4), CelD, 880 aa;
fasta scores: opt: 544 z-score: 555.0 E(): 1.5e-23; 42.8%
identity in 215 aa overlap. Contains Pfam matches to
entries PF00553 CBD 2, cellulose binding domain and
PF00041 fn3, fibronectin type III domain. Overlaps with
Streptomyces coelicolor str73.01c, leni. >327 aa;
C-terminus similar to many e.g. MANA, gp121, p5129 mannan
endo-1,4-beta-mannosidase precursor (EC 3.2.1.78) (363
aa), fasta scores: opt: 1189 z-score: 1332.7 E(): 0, 61.1%
identity in 288 aa overlap. Contains N-terminal signal
sequence, and Pfam match to entry PF00150 cellulase,
cellulase (glycosyl hydrolase family 5)⁹
/codon_start=1
/cdon_start=1
/transl_table=11
/product="putative secreted beta-mannosidase"
/protein_id="CAD55266.1"
/db_xref="GI:24413729"
/db_xref="GOA:O8CK53"
/db_xref="UniProt/TREMBL:O8CK53"
/translation="MRPARRPTPTLPRLITGLAALIGLVVAGALCRGAAASPAA
RSAGVPSAGADAGLITGDRLENGNDPVMRGVNAHTVPGSTSLADVAKLAGAN
SVRVSLDGRHRSNGPADVAVALI EOCKANRLICVLEVHDTTGYAEDAAAGTLHDAD
TWIGLKDVLACQEDYVIVNIGNEPMPGTEPTVAAYKULAAAGLQHTIVDAP
NMGGDQGVMAANASVYDADPTGNLIFSIHMYEPTAOITDYLYNAFVAEPILII
GEGGPDADGVDPDDEPTDMATAPBLRLGYLAMSNGNTDPLDIALDPDRSLGWER
RIRHGVNGLAQTSEBATTYGGADPDDEAPAPAPASPTANTDSYTLGVAATDDTA
VSGYDVAVLGGTSTVAASSTTTATVGLTAATYTAFAVARADAGRSKRSATVEY
TTDGGTTPGAGSVGRVGMFGPQGLAVRNTGSAIAGPMTLAPAFADGQVVTNM
MGTTATPDAGAVASVAPASVATITIPAGTVITIGTAPARKDNTATAPAAFLNAGVATP"
complement(3703..4005)
/gene="SC00554"

misc_feature

Query Match 94.1%; Score 16; DB 1; Length 314100;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2 CCGCGCGGTCGGGGAC 17
Db 82543 CCGCGCGGTCGGGGAC 82558

RESULT 12
AF133913 356 bp mRNA linear ROD 04-NOV-1999
DEFINITION Mus musculus ARL-6 interacting protein-6 (Aip-6) mRNA, partial cds.
ACCESSION AF133913
VERSION AF133913.1 GI:4927207
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Ingley, E., Williams, J.H., Walker, C.E., Teal, S., Colley, S.,
Sayer, M.S., Tibbrook, P.A., Sarna, M., Beaumont, J.G. and Klinken, S.P.
TITLE A novel ADP-ribosylation like factor (ARL-6), interacts with the
protein-conducting channel SSG1beta subunit
JOURNAL FEBS Lett. 459 (1), 69-74 (1999)
MEDLINE 99439712
PUBMED 10508919
REFERENCE 2 (bases 1 to 356)
AUTHORS Ingley, E., Teal, S. and Klinken, S.P.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-1999) Biochemistry, University of Western
Australia, Rear 50 Murray St., Perth, WA 6000, Australia
FEATURES
location/Qualifiers
1..356
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone_lib="yeast two-hybrid EML C.1"

gene

CDS

<1..>356
/gene="Aip-6"
CDS
1..>356
/gene="Aip-6"
/codon_start=1
/product="ARL-6 interacting protein-6"
/protein_id="AA033051.1"
/db_xref="GI:4927208"
/translation="MRPARRPTPTLPRLITGLAALIGLVVAGALCRGAAASPAA
RSAGVPSAGADAGLITGDRLENGNDPVMRGVNAHTVPGSTSLADVAKLAGAN
SVRVSLDGRHRSNGPADVAVALI EOCKANRLICVLEVHDTTGYAEDAAAGTLHDAD
TWIGLKDVLACQEDYVIVNIGNEPMPGTEPTVAAYKULAAAGLQHTIVDAP
NMGGDQGVMAANASVYDADPTGNLIFSIHMYEPTAOITDYLYNAFVAEPILII
GEGGPDADGVDPDDEPTDMATAPBLRLGYLAMSNGNTDPLDIALDPDRSLGWER
RIRHGVNGLAQTSEBATTYGGADPDDEAPAPAPASPTANTDSYTLGVAATDDTA
VSGYDVAVLGGTSTVAASSTTTATVGLTAATYTAFAVARADAGRSKRSATVEY
TTDGGTTPGAGSVGRVGMFGPQGLAVRNTGSAIAGPMTLAPAFADGQVVTNM
MGTTATPDAGAVASVAPASVATITIPAGTVITIGTAPARKDNTATAPAAFLNAGVATP"
SILCSLFAVLARFL"

ORIGIN

Query Match 90.6%; Score 15.4; DB 10; Length 356;
Best Local Similarity 94.1%; Pred. No. 1.5e+04;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TCCGCGGTCGGGGAC 17
Db 197 TCCGCGGTCGGGGAC 213

RESULT 13

LOCUS

PVU491803 795 bp mRNA linear INV 08-SEP-2002
DEFINITION Patella vulgata partial mRNA for wnt2 protein.
ACCESSION AJ491803
VERSION AJ491803.1 GI:21685397
KEYWORDS wnt2 protein.
SOURCE Patella vulgata (common limpet)
ORGANISM Eukaryota; Metazoa; Mollusca; Gastropoda; Bogastropoda; Docoglossa;
Patellina; Patellogoda; Patellicidae; Patella.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

ORIGIN

Query Match 90.6%; Score 15.4; DB 3; Length 795;
Best Local Similarity 94.1%; Pred. No. 1.3e+04;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCGCCGGGTGCGGAGC 17
Db 371 TCCGCCGGGTGCGGAGC 387

RESULT 14
RATPGP 1034 bp mRNA linear ROD 20-JAN-2000
LOCUS Rattus norvegicus mRNA for plasma glutathione peroxidase precursor,
DEFINITION complete cds.
ACCESSION D00680
VERSION D00680.1 GI:220843
KEYWORDS plasma glutathione peroxidase precursor; plasma glutathione peroxidase; seleno-protein.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 1034)
AUTHORS Yoshimura,S., Watanabe,K., Suenizu,H., Onozawa,T., Mizoguchi,J., Tsuda,K., Hata,H. and Moriuchi,T.
TITLE Tissue specific expression of the plasma glutathione peroxidase gene in rat kidney
JOURNAL J. Biochem. 109 (6), 918-923 (1991)
MEDLINE 92041746
PUBMED 1939013

COMMENT These data kindly submitted in computer readable form by: Shinichi Yoshimura
Dept. Cell. Biol.
Tokai University
School of Medicine
Bonseldai
Isehara
Kanagawa 259-11
Japan
Phone: 0463-93-1121 x2583
Fax: 0463-93-1121 x2583
Codon at position 350 to 352 is translated as selenocysteine.
Location/Qualifiers
1..1034
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Wistar"
/db_xref="taxon:10116"
/clone="pRGPx-3"
/tissue_type="placenta"
134..814
/EC_number="1.11.1.9"
/note="selenocysteine"
/codon_start=1
/transl_except=(pos:350..352,aa:Sec)
/product="Plasma glutathione peroxidase precursor"
/protein_id="BA00587.2"
/db_xref="GI:6723180"
/translation="MSRTLRASCLSLLAGVPVPRGQSEKTKDCHGMSGTIEYV
ALTDGEYVIRPKQYAKYILPVNAYUGLDPQLELALOEELGPRGLVIGPCN
OPGRQEPENSEILPSLKYRPGGFVNFQLEKGDVNGEGRFTPLKNSCPRA
ELGSPGLFWEPMKIHDIRNFEKFLVPGDIPIMRWYHRTVSNVMDILSYMRQ
AALGARGR"
134..190
191..811
/product="plasma glutathione peroxidase"

ORIGIN
Query Match 90.6%; Score 15.4; DB 10; Length 1034;
Best Local Similarity 94.1%; Pzed. No. 1.2e+04;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCGCCGGGTGCGGAGC 17
Db 190 TCCGCCGGGTGCGGAGC 206

RESULT 15
AK111058
LOCUS Oryza sativa (japonica cultivar-group) cDNA clone:002-175-D04, full
DEFINITION insert sequence.
ACCESSION AK111058
VERSION AK111058.1 GI:32996267
KEYWORDS FLI_CDNA; oligo capping.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehhartrioideae; Oryzaceae; Oryza.

REFERENCE 1
AUTHORS The Rice Full-Length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-Length cDNA Project Team;
Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
Kihimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
Ohtsuki,K., Shishiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Ii,C.,
Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Nishikura,J.,
Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN,
Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Oosato,N., Ota,Y.,
Saito,R., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T.,
Yoshino,M. and Hayashizaki,Y.
TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from
Japonica rice
JOURNAL Science 301 (5631), 376-379 (2003)
MEDLINE 12869764
PUBMED 12869764

2 (bases 1 to 1109)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoaka,T.,
Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Ii,C., Lu,M.,
Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,
Nanaki,T., Narikawa,R., Nishikura,J., Nishi,K., Nomura,K.,
Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
Oosato,N., Ota,Y., Otsu,Y., Ryu,R., Satoh,K., Sakai,C., Sakai,K.,
Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
Shingawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S.,
Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
Tagawa,A., Takahashi,F., Takaku-Ikashira,S., Tanaka,T., Tomaru,A.,
Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
Yoshimura,A.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail:skkuchiemias.affrc.go.jp.
Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT This clone is one of the 28k full-length cDNA clones from japonica
rice.
URL : <http://cdna01.dna.affrc.go.jp/cDNA/>
NIAS Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K.,
Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Nanaki,T.,
Ohneda,E., Yahagi,W., Suzuki,K., Ii,C., Ohtsuki,K., Shishiki,T. and
Yamamoto,M.

FAIS Genome Sequencing & Analysis Group: Otsu, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamura, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Maeda, H., Mura, U., Mizuno, K., Naitaka, R., Naitaka, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tanoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.
Genome Exploration Research Group in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Iida, J., Imamura, K., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishii, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yaenishi, A., and Hayashizaki, Y.

FEATURES

SOURCE

1. 1109
location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultiyar="Nipponbare"
/db_xref="taxon:3947"
/clone="002-175-D04"

ORIGIN

Query Match 90.6%; Score 15.4; DB 8; Length 1109;
Best Local Similarity 94.1%; Pred. No. 1.2e+04;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TCCGCCGGGTCGGGGAC 17
|||
Db 101 TCCGCCGGGTCGGGGAC 117

Search completed: November 7, 2005, 06:41:49
Job time : 676.064 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:04:21 ; Search time 196.043 Seconds
(without alignments)
513.336 Million cell updates/sec

Title: US-10-777-131A-2

Perfect score: 17
Sequence: 1 tccgcgcggcgcggggac 17

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

1: Geneseqn1980s:.*
2: Geneseqn1990s:.*
3: Geneseqn2000s:.*
4: Geneseqn2001as:.*
5: Geneseqn2001bs:.*
6: Geneseqn2002as:.*
7: Geneseqn2002bs:.*
8: Geneseqn2003as:.*
9: Geneseqn2003bs:.*
10: Geneseqn2003cs:.*
11: Geneseqn2003ds:.*
12: Geneseqn2004as:.*
13: Geneseqn2004bs:.*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	100.0	17	8	ABZ79928 Mycobacte
2	17	100.0	823	8	ABZ79951 Mycobacte
3	17	100.0	3002	13	ADST5303 Mycobacte
4	17	100.0	110000	4	AA199682_12
5	17	100.0	110000	4	AA199683_12
6	15.4	90.6	205	10	ADDD34800
7	15.4	90.6	337	12	ADH00311
8	15.4	90.6	707	3	AAA90423
9	15.4	90.6	879	8	ACA23119
10	15.4	90.6	1034	12	ADP72472
11	15.4	90.6	1194	2	AAT58555
12	15.4	90.6	1368	8	ACA25382
13	15.4	90.6	1415	12	ADJ75949
14	15.4	90.6	1528	2	AA277483
15	15.4	90.6	4496	2	AAT58553
16	15	88.2	144	6	ABQ90649
17	15	88.2	1167	3	AAC55858
18	15	88.2	1167	10	ADJ10267
19	15	88.2	1362	6	ABQ90990
20	15	88.2	2969	6	ABZ11506

C	21	15	88.2	2969	12	ADM44024	Adm44024 Novel hum
	22	15	88.2	3765	3	AAC55782	Aac55782 Nucleotid
	23	15	88.2	3765	10	ADJ10194	Adj10194 S. lavend
C	24	15	88.2	30156	12	ADK71085	Adk71085 Human MP2
C	25	15	88.2	53500	3	AAC55842	Aac55842 Complete
C	26	15	88.2	53500	10	ADJ10261	Adj10261 S. lavend
	27	15	88.2	110000	4	AA199682_12	Continuation (13 o
	28	15	88.2	110000	4	AA199682_12	Continuation (14 o
	29	15	88.2	110000	4	AA199683_12	Continuation (13 o
	30	15	88.2	110000	4	AA199683_13	Continuation (14 o
	31	15	88.2	250	4	AA136528	AA136528 Human mus
	32	14.4	84.7	250	8	ABX59516	Abx59516 CDNA enc
	33	14.4	84.7	250	12	ADJ30266	Adj30266 Human mus
	34	14.4	84.7	367	9	ACT13330	Act13330 DNA clone
	35	14.4	84.7	392	9	ACT13323	Act13323 DNA clone
C	36	14.4	84.7	464	9	ACH17641	Ach17641 Human adu
	37	14.4	84.7	475	9	ACT13258	Act13258 DNA clone
	38	14.4	84.7	499	9	ACT13288	Act13288 DNA clone
	39	14.4	84.7	501	9	ACT13255	Act13255 DNA clone
	40	14.4	84.7	525	9	ACT13279	Act13279 DNA clone
	41	14.4	84.7	540	9	ACT13232	Act13232 DNA clone
	42	14.4	84.7	540	9	ACT13217	Act13217 DNA clone
	43	14.4	84.7	545	9	ACT13348	Act13348 DNA clone
	44	14.4	84.7	549	9	ACT13239	Act13239 DNA clone
	45	14.4	84.7	558	9	ACT13227	Act13227 DNA clone

ALIGNMENTS

RESULT 1	ABZ79928	standard; DNA; 17 BP.
ID	ABZ79928	
XX	ABZ79928;	
AC	19-MAY-2003	(first entry)
DT		
XX		
DE	Mycobacterium tuberculosis mutT2-2 PCR primer SEQ ID NO:2.	
XX		
KW	Mycobacterium tuberculosis; mutT2; alkA; ogt; Rv3908; mutY; Rv3909;	
KM	detection; multidrug resistance; multiple drug resistance; MDR;	
XX	infection; PCR primer; ss.	
OS	Mycobacterium tuberculosis.	
XX	Synthetic.	
PN	WO2003016562-A2.	
XX		
PD	27-FEB-2003.	
XX		
PF	14-AUG-2002; 2002WO-EP009679.	
XX		
PR	14-AUG-2001; 2001US-0311824P.	
PR	21-AUG-2001; 2001US-0313523P.	
XX		
PA	(INST) INST PASTEUR.	
XX		
PI	Glcquel B;	
XX		
DR	WPI, 2003-256711/25.	
XX		
PT	Predicting the epidemic character of a Mycobacterium tuberculosis isolate	
PT	and/or the acquisition of multiple drug resistance (MDR) by the isolate	
PT	by detecting an alteration in the DNA repair system of the isolate.	
XX		
PS	Claim 32; Page 16; 83pp; English.	
XX		
CC	The present invention describes a method for predicting the epidemic	
CC	character of a Mycobacterium tuberculosis isolate and/or a selective	
CC	advantage to be maintained in the host and/or the acquisition of multiple	
CC	drug resistance (MDR) by the isolate comprising detecting an alteration	
CC	in the DNA repair system of the isolate. Also described: (1) detecting a	

CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a
CC polymnucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)
CC an Escherichia coli strain containing the plasmid pMYC2501; and (5)
CC detecting in a patient infected by Mycobacterium tuberculosis a higher
CC risk of being unable to eliminate the bacillus or of developing MDR
CC tuberculosis. The method is useful for predicting the epidemic character
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to
CC be maintained in the host and/or the acquisition of MDR by the isolate.
CC The present sequence represents a PCR primer for M. tuberculosis mutT2,
CC which is used in the exemplification of the present invention
CC
CC Sequence 17 BP; 1 A; 6 C; 8 G; 2 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 17; DB 8; Length 17;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TCCGCCGGGTGCGGAC 17
Db 1 TCCGCCGGGTGCGGAC 17
RESULT 2
ABZ79951/c
ID ABZ79951 standard; DNA; 823 BP.
XX
XX ABZ79951;
XX
XX 19-MAY-2003 (first entry)
XX
XX Mycobacterium tuberculosis mutT2 nucleotide sequence SEQ ID NO:19.
XX
XX Mycobacterium tuberculosis; mutT2; alka; ogt; RV3908; mutY; RV3909;
XX detection; multidrug resistance; multiple drug resistance; MDR;
XX infection; gene; ds.
XX
XX Mycobacterium tuberculosis.
XX
XX WO2003016562-A2.
XX
XX 27-FEB-2003.
XX
XX 14-AUG-2002; 2002WO-EP009679.
XX
XX 14-AUG-2001; 2001US-0311824P.
XX PR 21-AUG-2001; 2001US-0313523P.
XX
XX (INSP) INST PASTEUR.
XX
XX Glcque1 B;
XX
XX WPI; 2003-256711/25.
XX
XX Predicting the epidemic character of a Mycobacterium tuberculosis isolate
XX and/or the acquisition of multiple drug resistance (MDR) by the isolate
XX by detecting an alteration in the DNA repair system of the isolate.
XX
XX Disclosure; Fig 4A; 83pp; English.
XX
XX The present invention describes a method for predicting the epidemic
XX character of a Mycobacterium tuberculosis isolate and/or a selective
XX advantage to be maintained in the host and/or the acquisition of multiple
XX drug resistance (MDR) by the isolate comprising detecting an alteration
XX in the DNA repair system of the isolate. Also described: (1) detecting a
XX Mycobacterium tuberculosis strain with a MDR phenotype; (2) a
XX polymnucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)
XX an Escherichia coli strain containing the plasmid pMYC2501; and (5)
XX detecting in a patient infected by Mycobacterium tuberculosis a higher
XX risk of being unable to eliminate the bacillus or of developing MDR
XX tuberculosis. The method is useful for predicting the epidemic character
XX of a Mycobacterium tuberculosis isolate and/or a selective advantage to
XX be maintained in the host and/or the acquisition of MDR by the isolate.
XX The present sequence represents a M. tuberculosis mutT2 nucleotide

CC sequence, which is used in the exemplification of the present invention
XX
XX Sequence 823 BP; 144 A; 253 C; 277 G; 149 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 17; DB 8; Length 823;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TCCGCCGGGTGCGGAC 17
Db 820 TCCGCCGGGTGCGGAC 804
RESULT 3
ADS75303/c
ID ADS75303 standard; DNA; 3002 BP.
XX
XX ADS75303;
XX
XX 16-DEC-2004 (first entry)
XX
XX Mycobacterium tuberculosis nargHJI nitrate reductase operon Segid 1.
XX
XX nargHJI nitrate reductase operon; single nucleotide polymorphism; SNP;
XX detection method; ds.
XX
XX Mycobacterium tuberculosis.
XX
XX Key Location/Qualifiers
XX variation replace(1500,C)
XX /tag= a
XX /standard_name= "Single nucleotide polymorphism"
XX misc_feature 1717..1719
XX /tag= b
XX /note= "GNG start codon"
XX
XX WO2004083459-A1.
XX
XX 30-SEP-2004.
XX
XX 19-MAR-2004; 2004WO-EP002911.
XX
XX 20-MAR-2003; 2003DE-01013791.
XX
XX (ARTU-) ARTUS GBS MOLBKLARBIOLGISCHE DIAGNOSTI.
XX
XX Bange F;
XX
XX WPI; 2004-728487/71.
XX
XX Specific detection of Mycobacterium tuberculosis, useful for differential
XX diagnosis of infection, by detecting specific polymorphism in the
XX promoter of the nitrate reductase operon.
XX
XX Claim 1; SEQ ID NO 1; 46pp; German.
XX
XX This invention relates to a novel method for the detection of
XX Mycobacterium tuberculosis (M. tuberculosis) occurring in a biological
XX sample. Specifically, it refers to the amplification of the nargHJI
XX nitrate reductase operon, including position -215 in the 5' to 3' reading
XX direction that occurs upstream of the GNG start codon of the operon. The
XX present invention describes detecting an M. tuberculosis specific
XX polymorphism at position -215. The polymorphism (C-215T) is detected by
XX hybridisation to specific probes such that M. tuberculosis can be
XX differentiated from M. bovis, M. bovis BCG, M. africanum and M. microti
XX by analysis of the melting curve. The method is used to detect M.
XX tuberculosis infection, especially to differentiate it from other species
XX of the M. tuberculosis complex, and furthermore provides rapid and
XX specific detection, and differentiation of M. tuberculosis. This
XX polymnucleotide sequence is the M. tuberculosis nargHJI nitrate reductase
XX operon containing the -215 SNP at position 1500 in this sequence of the
XX invention.

SQ Sequence 3002 BP; 483 A; 923 C; 1013 G; 583 T; 0 U; 0 Other;

Query Match 100.0%; Score 17; DB 13; Length 3002;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGCCGGGTCCGGGAC 17
Db 1603 TCCGCCGGGTCCGGGAC 1587

RESULT 4

AA199682_12/c

Continuation (13 of 45) of AA199682 from base 1200001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AA199682 Accession AA199682

WP	Fragment Name	Begin	End
WP	AA199682_00	1	110000
WP	AA199682_01	100001	210000
WP	AA199682_02	200001	310000
WP	AA199682_03	300001	410000
WP	AA199682_04	400001	510000
WP	AA199682_05	500001	610000
WP	AA199682_06	600001	710000
WP	AA199682_07	700001	810000
WP	AA199682_08	800001	910000
WP	AA199682_09	900001	1010000
WP	AA199682_10	1000001	1110000
WP	AA199682_11	1100001	1210000
WP	AA199682_12	1200001	1310000
WP	AA199682_13	1300001	1410000
WP	AA199682_14	1400001	1510000
WP	AA199682_15	1500001	1610000
WP	AA199682_16	1600001	1710000
WP	AA199682_17	1700001	1810000
WP	AA199682_18	1800001	1910000
WP	AA199682_19	1900001	2010000
WP	AA199682_20	2000001	2110000
WP	AA199682_21	2100001	2210000
WP	AA199682_22	2200001	2310000
WP	AA199682_23	2300001	2410000
WP	AA199682_24	2400001	2510000
WP	AA199682_25	2500001	2610000
WP	AA199682_26	2600001	2710000
WP	AA199682_27	2700001	2810000
WP	AA199682_28	2800001	2910000
WP	AA199682_29	2900001	3010000
WP	AA199682_30	3000001	3110000
WP	AA199682_31	3100001	3210000
WP	AA199682_32	3200001	3310000
WP	AA199682_33	3300001	3410000
WP	AA199682_34	3400001	3510000
WP	AA199682_35	3500001	3610000
WP	AA199682_36	3600001	3710000
WP	AA199682_37	3700001	3810000
WP	AA199682_38	3800001	3910000
WP	AA199682_39	3900001	4010000
WP	AA199682_40	4000001	4110000
WP	AA199682_41	4100001	4210000
WP	AA199682_42	4200001	4310000
WP	AA199682_43	4300001	4410000
WP	AA199682_44	4400001	4411529

Query Match 100.0%; Score 17; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGCCGGGTCCGGGAC 17
Db 87212 TCCGCCGGGTCCGGGAC 87196

RESULT 5

AA199683_12/c

Continuation (13 of 44) of AA199683 from base 1200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AA199683 Accession AA199683

WP	Fragment Name	Begin	End
WP	AA199683_00	1	110000
WP	AA199683_01	100001	210000
WP	AA199683_02	200001	310000
WP	AA199683_03	300001	410000
WP	AA199683_04	400001	510000
WP	AA199683_05	500001	610000
WP	AA199683_06	600001	710000
WP	AA199683_07	700001	810000
WP	AA199683_08	800001	910000
WP	AA199683_09	900001	1010000
WP	AA199683_10	1000001	1110000
WP	AA199683_11	1100001	1210000
WP	AA199683_12	1200001	1310000
WP	AA199683_13	1300001	1410000
WP	AA199683_14	1400001	1510000
WP	AA199683_15	1500001	1610000
WP	AA199683_16	1600001	1710000
WP	AA199683_17	1700001	1810000
WP	AA199683_18	1800001	1910000
WP	AA199683_19	1900001	2010000
WP	AA199683_20	2000001	2110000
WP	AA199683_21	2100001	2210000
WP	AA199683_22	2200001	2310000
WP	AA199683_23	2300001	2410000
WP	AA199683_24	2400001	2510000
WP	AA199683_25	2500001	2610000
WP	AA199683_26	2600001	2710000
WP	AA199683_27	2700001	2810000
WP	AA199683_28	2800001	2910000
WP	AA199683_29	2900001	3010000
WP	AA199683_30	3000001	3110000
WP	AA199683_31	3100001	3210000
WP	AA199683_32	3200001	3310000
WP	AA199683_33	3300001	3410000
WP	AA199683_34	3400001	3510000
WP	AA199683_35	3500001	3610000
WP	AA199683_36	3600001	3710000
WP	AA199683_37	3700001	3810000
WP	AA199683_38	3800001	3910000
WP	AA199683_39	3900001	4010000
WP	AA199683_40	4000001	4110000
WP	AA199683_41	4100001	4210000
WP	AA199683_42	4200001	4310000
WP	AA199683_43	4300001	4403765

Query Match 100.0%; Score 17; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGCCGGGTCCGGGAC 17
Db 86681 TCCGCCGGGTCCGGGAC 86665

RESULT 6

ADD34800

ID ADD34800 standard; DNA; 205 BP.
AC ADD34800;

DT 15-JAN-2004 (first entry)

DE Mouse mitochondrial DNA sequence SEQ ID NO:2579.

XX ds; mouse; array; mitochondrial; hybridisation; energy-metabolism;
KW mitochondrial disease; oxidative phosphorylation dysfunction;
XX oxidative stress; apoptosis; aging.
XX Mus musculus.
XX

```
PN WO2003020220-A2.
XX
XX 13-MAR-2003.
PD
XX
XX 30-AUG-2002; 2002WO-US027886.
PF
XX 30-AUG-2001; 2001US-0316323P.
XX
XX 31-AUG-2001; 2001CA-02356540.
PR
XX
XX (UYEM-) UNIV EMORY.
PA
XX
XX Wallace DC, Levy S, Kerstam K, Procaccio V;
PI
XX WPI; 2003-300821/29.
DR
XX
XX
XX Array containing probes for genes involved in mitochondrial biology,
PT useful for determining mitochondrial biology gene expression profiles for
PT use in diagnosing pathologies and identifying biochemical pathways.
XX
XX Claim 2; SEQ ID NO 2579; 201pp; English.
PS
XX
XX The invention relates to a novel array comprising at least two isolated
XX nucleotide molecules, each molecule having a sequence capable of uniquely
XX hybridizing to a nucleic acid molecule which is an expression product of
XX a gene involved in mitochondrial biology. The array comprises two or more
XX isolated nucleic acid molecules or spots, each molecule having a sequence
XX chosen from sequence of 994 human probes and 2046 mouse probes. An array
XX of the invention is useful for determining an expression profile of a
XX mouse or human sample containing nucleic acid, by contacting the array
XX with the sample under conditions allowing selective hybridisation, and
XX measuring hybridisation of nucleic acid in the sample to the array to
XX produce an expression profile. The array is also useful for determining
XX an expression profile of a first labelled sample containing nucleic acid
XX relative to a second, differently labelled sample containing nucleic
XX acid. The second sample is a reference or a standard. An array is useful
XX for determining an expression profile diagnostic of an energy-metabolism-
XX related physiological condition. An array of the invention is useful for
XX determining mitochondrial biology gene expression profiles of organisms,
XX such as human, mice and closely related species, tissue and organs of
XX such organisms, which are useful for determining expression profiles
XX diagnostic of energy metabolism-related physiological conditions,
XX diagnosing such physiological conditions, identifying biochemical
XX pathways, genes, and mutations involved in such physiological conditions,
XX identifying therapeutic agents useful for preventing and/or treating such
XX physiological conditions, evaluating and/or monitoring the efficacy of
XX such therapies, and creating and identifying animal models of human
XX energy metabolism-related physiological conditions. An array is also
XX useful for defining expression signatures or profiles for mitochondrial
XX diseases, as well as distinguishing clinical disorders that result from
XX oxidative phosphorylation (OXPHOS) dysfunction, oxidative stress,
XX apoptosis and aging. An array of the invention contains probes of genes
XX not previously recognised to participate in mitochondrial biology. The
XX sequences shown in ADD33224-ADD35260 represent murine mitochondrial DNA
XX clones used to make the probes of the invention. Some sequences are not
XX present, these are SEQ ID NO's 295, 1174, 1213, 1700, 1728, 1730, 1905,
XX 1906, 2408 and 2643.
XX
XX Sequence 205 BP; 40 A; 66 C; 58 G; 41 T; 0 U; 0 Other;
SQ
XX
XX Query Match 90.6%; Score 15.4; DB 10; Length 205;
XX Best Local Similarity 94.1%; Pred. No. 1.4e+03;
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TCCGCGGGGTGCGGAC 17
XX ||||||| |||||||
DB 85 TCCGCGGGGCGCGGAC 101
XX
XX RESULT 7
XX ADH00311
XX ID ADH00311 standard; DNA; 337 BP.
XX
XX AC ADH00311;
```

```
XX
XX 11-MAR-2004 (first entry)
DT
XX
XX Kidney disease-associated gene-related mouse DNA sequence SeqID1440.
DE
XX
XX Kidney disease; proximal tubule; nephrotropic; chronic renal failure;
XX kidney disease; proximal tubule; nephrotropic; chronic renal failure;
XX diabetic nephropathy; mouse; murine; ds.
XX
XX Mus sp.
OS
XX
XX WO2003091427-A1.
XX
XX 06-NOV-2003.
PD
XX
XX
XX 23-APR-2003; 2003WO-JP005137.
XX
XX 24-APR-2002; 2002JP-00122253.
XX
XX (KANS-) KANSAI TECHNOLOGY LICENSING ORG CO LTD.
XX
XX Takenaka M, Imai E, Okubo K;
PI
XX
XX WPI; 2004-011769/01.
DR
XX
XX Kidney disease-associated genes expressed in proximal tubule, applicable
XX in diagnosis of, in drug development and in devising appropriate strategy
XX in treatment of e.g. chronic renal failure and diabetic nephropathy.
PT
XX
XX Claim 1; SEQ ID NO 1440; 770pp; Japanese.
PS
XX
XX This invention relates to novel kidney disease-associated genes expressed
XX in proximal tubules. The invention may be useful for the development of
XX compounds with a nephrotropic activity. The genes are applicable in the
XX diagnosis of, in drug development and in devising appropriate strategy in
XX the treatment of (for example) chronic renal failure and diabetic
XX nephropathy. The present sequence is that of a mouse DNA sequence which is
XX related to the invention.
XX
XX Sequence 337 BP; 83 A; 87 C; 85 G; 74 T; 0 U; 8 Other;
SQ
XX
XX Query Match 90.6%; Score 15.4; DB 12; Length 337;
XX Best Local Similarity 94.1%; Pred. No. 1.3e+03;
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TCCGCGGGGTGCGGAC 17
XX ||||||| |||||||
DB 49 TCCGCGGGGCGCGGAC 65
XX
XX RESULT 8
XX AAA90423
XX ID AAA90423 standard; CDNA; 707 BP.
XX
XX AAA90423;
AC
XX
XX 10-JAN-2001 (first entry)
DT
XX
XX Human plasma glutathione peroxidase H (GPxH) cDNA.
XX
XX Plasma glutathione peroxidase H; GPxH; human; recombinant production;
XX ss.
XX
XX Homo sapiens.
OS
XX
XX Key
XX CDS
XX
XX Location/Qualifiers
XX 16..696
XX
XX /tag= a
XX /product= "Human GPx4"
XX /transcript= (pos:232..234, aa:Xaa)
XX /note= "These nucleotides represent an in-frame stop
XX codon (TGA); Xaa = unknown"
XX
XX CN1256312-A.
```



```
XX 14-JUN-2000.
PD 29-OCT-1998; 98CN-00121973.
XX
XX 29-OCT-1998; 98CN-00121973.
XX
XX (UYFU-) UNIV FUDAN.
XX
XX Yu L, Tu Q, Fu Q;
XX
XX WPI; 2000-533631/49.
XX
XX P-PSDB; AAB22828.
XX
XX New human glutathione peroxidase and its code sequence, preparation and
XX use.
XX
XX Claim 1; Page 20; 26pp; Chinese.
XX
XX This sequence represents cDNA encoding human plasma glutathione
XX peroxidase H (GPxH). The invention relates to this novel human
XX glutathione peroxidase, nucleic acid encoding it, and to recombinant
XX production of human GPxH. The invention also encompasses applications
XX for human GPxH.
XX
XX Sequence 707 BP; 165 A; 191 C; 192 G; 159 T; 0 U; 0 Other;
SQ
XX
XX Query Match 90.6%; Score 15.4; DB 3; Length 707;
XX Best Local Similarity 94.1%; Pred. No. 1.2e+03;
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TCCGCCGGGTCGGGAC 17
Db 72 TCCGCCGGGTCGGGAC 88
XX
XX RESULT 9
XX ACA23319/c
XX ID ACA23319 standard; DNA; 879 BP.
XX
XX ACA23319;
XX
XX 19-JUN-2003 (first entry)
XX
XX Prokaryotic essential gene #4976.
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
XX Borrelia cepacia.
XX
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0342923P.
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
XX
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX
XX P-PSDB; ABU19449.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
```

```
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 11189; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is one of the target
XX not form part of the printed specification, but was obtained in
XX electronic format directly from WIP0 at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 879 BP; 178 A; 300 C; 265 G; 136 T; 0 U; 0 Other;
SQ
XX
XX Query Match 90.6%; Score 15.4; DB 8; Length 879;
XX Best Local Similarity 94.1%; Pred. No. 1.1e+03;
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TCCGCCGGGTCGGGAC 17
Db 53 TCCGCCGGGTCGGGAC 37
XX
XX RESULT 10
XX ADP72472
XX ID ADP72472 standard; DNA; 1034 BP.
XX
XX ADP72472;
XX
XX 26-AUG-2004 (first entry)
XX
XX Renal toxin progression gene marker #1061.
XX
XX ds; toxic effect; gene expression profile; kidney tissue;
XX differential gene expression; toxicity progression; toxicity marker;
XX drug screening; toxicity assay; kidney pathology; nephritis;
XX kidney necrosis; glomerular injury; tubular injury;
XX focal segmental glomerulosclerosis.
XX
XX Rattus norvegicus.
XX
XX WO2004048598-A2.
XX
XX 10-JUN-2004.
XX
XX 24-NOV-2003; 2003WO-US037556.
XX
XX 22-NOV-2002; 2002US-00301856.
XX
```

PA (GENE-) GENE LOGIC INC.
XX Mendrick DL, Porter MW, Johnson KR, Castle A, Higgs B;
PI Elashoff M;
XX WPI; 2004-460771/43.
XX
XX Predicting (the progression of) a toxic effect of a compound, for
PT monitoring the progression of renal disease states, comprises preparing a
PT gene expression profile of a kidney tissue or cell sample exposed to the
PT compound.
XX
XX Claim 11; SEQ ID NO 1061; 266pp; English.
XX
XX The invention relates to a method of predicting (the progression of) a
XX toxic effect of a compound by preparing a gene expression profile of a
XX kidney tissue or cell sample exposed to the compound and comparing the
XX gene expression profile to a database, or detecting the level of gene(s)
XX expression in a tissue or cell sample exposed to the compound, where
XX differential gene expression compared to a control indicates a toxic
XX effect (toxicity progression). The method is useful for predicting (the
XX progression of) at least one toxic effect of a compound. The genes are
XX useful as toxicity markers in drug screening and toxicity assays. The
XX methods are useful for predicting the likelihood that a compound or test
XX agent will induce various specific kidney pathologies, such as nephritis,
XX kidney necrosis, glomerular and tubular injury, or focal segmental
XX glomerulosclerosis. The methods are useful for determining the similarity
XX of a toxic response to one or more individual compounds and for
XX predicting or elucidating the potential cellular pathways influenced,
XX induced or modulated by the compound or test agent. The kit is useful for
XX predicting or modelling the toxic response of a test compound, for
XX monitoring the progression of renal disease states, for identifying genes
XX that show promise as new drug targets and for screening known and newly
XX designed drugs. This sequence corresponds to a gene marker used in the
XX method of the invention. (Note: The sequence data for this patent did not
XX form part of the printed specification, but was obtained in electronic
XX format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences).
XX
XX Sequence 1034 BP; 247 A; 298 C; 277 G; 212 T; 0 U; 0 Other;
XX
XX
XX Query Match 90.6%; Score 15.4; DB 12; Length 1034;
XX Best Local Similarity 94.1%; Pred. No. 1.1e+03;
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 TCCGCCGGGTCGGGAC 17
XX |||||||
XX 190 TCCGCCGGGTCGGGAC 206
XX
XX
XX RESULT 11
XX AAT58555/c
XX ID AAT58555 standard; cDNA; 1194 BP.
XX
XX AAT58555;
XX
XX 02-APR-1997 (first entry)
XX
XX Streptomyces pristinaespiralis snbF gene.
XX
XX Streptogramin B; antibiotic; biosynthesis; pristinaamycin; virginiamycin;
XX pipecolic acid; cyclodeamination; papa; snbA; snbF; p1pA;
XX 3-hydroxypicolinic acid; hydroxylation; ds.
XX
XX Streptomyces pristinaespiralis.
XX
XX
XX Key Location/Qualifiers
XX mat_peptide 1..1194
XX FT /*tag= a
XX FT /product= "SnbF"
XX
XX MO9601901-A1.
XX
XX 25-JAN-1996.
XX
XX PD

XX
XX 04-JUL-1995; 95WO-FR000889.
XX
XX 08-JUL-1994; 94FR-00008478.
XX
XX (RHON) RHONE POULENC RORER SA.
XX
XX Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J,
XX Barriere J, Debussche L, Famechon A, Paris J, Dutruc-Rosset G;
XX WPI; 1996-097631/10.
XX
XX P-PSDB; AAM11585.
XX
XX New streptogramin B derivative, useful as antibiotics - produced by new
XX mutants of Streptomyces having altered genes for streptogramin B
XX biosynthesis.
XX
XX Claim 18; Page 113-114; 146pp; French.
XX
XX The papa gene of S.pristinaespiralis is involved in the biosynthesis of 4
XX -dimethylamino-L-phenylalanine (DMPAPA), a precursor for pristinaamycin
XX IA. Upstream of the papa gene, on the complementary strand, is the snbA
XX gene coding for 3-hydroxypicolinic acid-AMP ligase. The region between
XX these two genes was sequenced and two open reading frames were
XX identified. The first (p1pA) decodes to an amino acid sequence with
XX homology to ornithine cyclodeaminase from Agrobacterium tumefaciens. The
XX p1pA gene product is likely to catalyse the cyclo- deamination of lysine,
XX leading to production of pipecolic acid. Mutations in the p1pA gene were
XX shown to affect pipecolic acid synthesis but not the synthesis of 3-
XX hydroxypicolinic acid. The second open reading frame (snbF) could be
XX decoded to give a product with homology to hydroxylases of the cytochrome
XX P450 type. Disruption of the p1pA and snbF genes can be used to produce
XX strains of S.pristinaespiralis which are unable to produce the antibiotic
XX pristinaamycin I but which may be able to produce new, modified forms of
XX it
XX
XX Sequence 1194 BP; 189 A; 522 C; 321 G; 162 T; 0 U; 0 Other;
XX
XX
XX Query Match 90.6%; Score 15.4; DB 2; Length 1194;
XX Best Local Similarity 94.1%; Pred. No. 1.1e+03;
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 TCCGCCGGGTCGGGAC 17
XX |||||||
XX 500 TCCGCCGGGTCGGGAC 484
XX
XX
XX RESULT 12
XX ACA25382/c
XX ID ACA25382 standard; DNA; 1368 BP.
XX
XX ACA25382;
XX
XX 19-JUN-2003 (first entry)
XX
XX Prokaryotic essential gene #7039.
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
XX Burkholderia fungorum.
XX
XX WO20027183-A2.
XX
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX
XX PR

XX	(ELIT-)	ELITRA PHARM INC.
PA		
XX	Wang L,	Zamudio C, Malone C, Haasebeck R, Ohlsen KL, Zykkind JW;
XX	P1	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX	DR	WPI; 2003-029926/02.
XX	P-Psdb;	ABU21512.
PT	New antisense nucleic acids, useful for identifying proteins or screening	
PT	for homologous nucleic acids required for cellular proliferation to	
PT	isolate candidate molecules for rational drug discovery programs.	
XX		
PS	Claim 14; SEQ ID NO 13252; 1766bp; English.	
XX		
CC	The invention relates to an isolated nucleic acid comprising any one of	
CC	the 6213 antisense sequences given in the specification where expression	
CC	of the nucleic acid inhibits proliferation of a cell. Also included are:	
CC	(1) a vector comprising a promoter operably linked to the nucleic acid	
CC	encoding a polypeptide whose expression is inhibited by the antisense	
CC	nucleic acid; (2) a host cell containing the vector; (3) an isolated	
CC	polypeptide or its fragment whose expression is inhibited by the	
CC	antisense nucleic acid; (4) an antibody capable of specifically binding	
CC	the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular	
CC	proliferation or the activity of a gene in an operon required for	
CC	proliferation; (7) identifying a compound that influences the activity of	
CC	the gene product or that has an activity against a biological pathway	
CC	required for proliferation, or that inhibits cellular proliferation; (8)	
CC	identifying a gene required for cellular proliferation or the biological	
CC	pathway in which a proliferation-regulated gene or its gene product lies	
CC	or a gene on which the test compound that inhibits proliferation of an	
CC	organism acts; (9) manufacturing an antibiotic; (10) profiling a	
CC	compound's activity; (11) a culture comprising strains in which the gene	
CC	product is overexpressed or underexpressed; (12) determining the extent	
CC	to which each of the strains is present in a culture or collection of	
CC	strains; or (13) identifying the target of a compound that inhibits the	
CC	proliferation of an organism. The antisense nucleic acids are useful for	
CC	identifying proteins or screening for homologous nucleic acids required	
CC	for cellular proliferation to isolate candidate molecules for rational	
CC	drug discovery programs, or for screening homologous nucleic acids	
CC	required for proliferation in cells other than S. aureus, S. typhimurium,	
CC	K. pneumoniae or P. aeruginosa. The present sequence is one of the target	
CC	prokaryotic essential genes. Note: The sequence data for this patent did	
CC	not form part of the printed specification, but was obtained in	
CC	electronic format directly from WIPO at	
CC	ftp.wipo.int/pub/published_pct_sequences	
XX		
SQ	Sequence 1368 BP; 156 A; 439 C; 477 G; 296 T; 0 U; 0 Other;	
	Query Match	90.6%; Score 15.4; DB 8; Length 1368;
	Best Local Similarity	94.1%; Pred. No. 1.e+03;
	Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Oy	1 TCCGCGGGATCGGCAGAC 17	
Db	467 TCCGCCGGATCGGCAGAC 451	
RESULT 13		
ADJ75949		
ID	ADJ75949 standard; DNA; 1415 BP.	
XX		
AC	ADJ75949;	
XX		
DT	20-MAY-2004 (first entry)	
DE	Marker gene SEQ ID NO:1201.	
XX		
KM	bronchial asthma; chronic obstructive pulmonary disease;	
KM	respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;	
XX	gene therapy; marker gene; gene; ds.	
XX		
OS	Mus musculus	

XX	PM	EP1394274-A2.
XX	PD	03-MAR-2004.
XX	PF	04-AUG-2003; 2003EP-00254857.
XX	PR	06-AUG-2002; 2002JP-00229312.
XX	PR	20-MAR-2003; 2003JP-00077212.
XX	PA	(GENO-) GENOX RES INC.
XX	PI	Ohtani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuhara K;
XX	DR	WPI; 2004-193155/19.
XX	PT	Testing for bronchial asthma or chronic obstructive pulmonary disease by
XX	PT	comparing the expression level of a marker gene in a biological sample
XX	PT	from a subject with the expression level of the gene in a sample from a
XX	PT	healthy subject.
XX	PS	Example 11; SEQ ID NO 1201; 241pp; English.
XX	CC	The present invention describes a method of testing for bronchial asthma
XX	CC	or chronic obstructive pulmonary disease. The method comprises
XX	CC	determining the expression level of a marker gene in a biological sample
XX	CC	from a subject, comparing the expression level determined with the
XX	CC	expression level of the marker gene in a biological sample from a healthy
XX	CC	subject, and judging whether the subject has bronchial asthma or chronic
XX	CC	obstructive pulmonary disease. The marker gene comprises: (a) a group of
XX	CC	genes (S1) whose expression levels increase when respiratory epithelial
XX	CC	cells are stimulated with interleukin-13; or (b) a group of genes (S2)
XX	CC	whose expression levels decrease when respiratory epithelial cells are
XX	CC	stimulated with interleukin-13. Also described: (1) a reagent (I) for
XX	CC	testing for bronchial asthma or chronic obstructive pulmonary disease;
XX	CC	(2) a kit for screening for a candidate compound for a therapeutic agent
XX	CC	to treat bronchial asthma or chronic obstructive pulmonary disease; (3)
XX	CC	an animal model for bronchial asthma or chronic obstructive pulmonary
XX	CC	disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a
XX	CC	method for producing an animal model for bronchial asthma or chronic
XX	CC	obstructive pulmonary disease; (6) a therapeutic agent for bronchial
XX	CC	asthma or chronic obstructive pulmonary disease, comprising the compound,
XX	CC	a marker gene or an antisense nucleic acid corresponding to a portion of
XX	CC	the marker gene, a ribozyme, a polynucleotide that suppresses the
XX	CC	expression of the gene through an RNAi effect or an antibody recognising
XX	CC	a protein encoded by a marker gene; and (7) a DNA chip for testing for
XX	CC	bronchial asthma or a chronic obstructive pulmonary disease, on which a
XX	CC	probe has been immobilised to assay a marker gene. (I) has respiratory
XX	CC	and antiasthmatic activities, and can be used in gene therapy. The method
XX	CC	is useful for testing for or screening for a therapeutic agent for
XX	CC	bronchial asthma or chronic obstructive pulmonary disease. The present
XX	CC	sequence is used in the exemplification of the present invention.
XX	SEQ	Sequence 1415 BP; 319 A; 437 C; 346 G; 313 T; 0 U; 0 Other;
XX	Query Match	90.6%; Score 15.4; DB 12; Length 1415;
XX	Best Local Similarity	94.1%; Pred. NO.1,1e+03;
XX	Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
XX	1 TCCGCCGGGATCGGGAC 17	
XX		
XX	Db 73 TCCGCCGGGATCGGGAC 89	
XX	RESULT 14	
XX	ID AA277483	
XX	AC AA277483	
XX	DT 10-APR-2000 (first entry)	
XX	DE Human ovarian tumor cDNA library derived EST fragment 34.	

XX Expressed sequence tag; EST; human; ovarian tumor; anticancer;
 KW gene therapy; treatment; se.
 XX Homo sapiens.
 XX DB19817557-A1.
 XX 21-OCT-1999.
 XX 09-APR-1998; 98DE-01017557.
 XX 09-APR-1998; 98DE-01017557.
 XX 09-APR-1998; 98DE-01017557.
 XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
 DR WPI; 1999-591920/51.
 DR P-PSDB; AAY76581, AAY76582.
 XX New nucleic acid sequences expressed in ovarian, and some other, cancer
 PT tissues, and derived polypeptides, for treatment of ovarian cancer and
 FT identification of therapeutic agents.
 XX
 XX Claim 3; Page 165; 310pp; German.
 XX This invention describes novel nucleic acid (cDNA) sequences (A) which
 CC have anticancer activity and are highly expressed in ovarian tumor tissue
 CC (and some also in testis and breast cancer tissue). The products of the
 CC invention can be used for gene therapy. (A) are used (i) for recombinant
 CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)
 CC are used (i) to identify agents suitable for treatment of ovarian cancer;
 CC (ii) directly for treating this form of cancer (including expression from
 CC gene therapy vectors) and (iii) for generation of specific antibodies.
 CC (A) are identified by assembling ESTs (expressed sequence tags) from a
 CC particular tissue type before comparison of expression patterns. This
 CC allows a significantly longer fragment of the gene to be revealed, so
 CC should reduce the number of failures associated with the fact that ESTs
 CC from different libraries may represent different parts of the same
 CC unknown gene, distorting the estimated frequency of occurrence in a
 CC particular tissue. AAZ77450-Z77572 represent the human ovarian tumor cDNA
 CC library derived EST fragments described in the method of the invention
 CC and encode the protein fragments represented in AAY76505-Y76638
 CC
 XX SQ Sequence 1528 BP; 365 A; 459 C; 373 G; 331 T; 0 U; 0 Other;
 Query Match 90.6%; Score 15.4; DB 2; Length 1528;
 Best Local Similarity 94.1%; Pred. No. 1.1e+03;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 TCCGCCGGGTCCGGGAC 17
 Db 157 TCCGCCGGGCGCGGAC 173
 RESULT 15
 AAT58553/C
 ID AAT58553 standard; cDNA; 4496 BP.
 XX AAT58553;
 XX 02-APR-1997 (first entry)
 XX Streptomyces pristinaespiralis snbA and papA intergenic region.
 XX Streptomyces pristinaespiralis; pristinamycin; virginiamycin;
 KW streptomycin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;
 KW pipercolic acid; cyclodeamination; papA; snbA; snbF; pipA;
 KW 3-hydroxypicolinic acid; hydroxylation; ds.
 XX Streptomyces pristinaespiralis.
 XX Key Location/Qualifiers

FT CDS complement(1..332)
 FT /*tag= a
 FT /product= "snbA"
 FT /note= "N-terminal coding region only, i.e. a partial
 FT open reading frame"
 FT CDS 607..1674
 FT /*tag= b
 FT /product= "pipA"
 FT 1800..2996
 FT /*tag= c
 FT /product= "snbF"
 FT 3018..4496
 FT /*tag= d
 FT /product= "papA"
 FT /note= "N-terminal coding region only, i.e. a partial
 FT open reading frame"
 XX
 XX W09601901-A1.
 XX 25-JAN-1996.
 XX 04-JUL-1995; 95MO-FR000889.
 XX 08-JUL-1994; 94FR-00008478.
 XX (RHON) RHONE POULENC RORER SA.
 XX Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J;
 PI Barriere J, Debussche L, Famechon A, Paris J, Dutruc-Rosset G;
 XX WPI; 1996-097631/10.
 DR P-PSDB; AAM11584, AAM11585.
 XX New streptogramin B derivative, useful as antibiotics - produced by new
 PT mutants of Streptomyces having altered genes for streptomycin B
 PT biosynthesis.
 XX
 XX Example 1; Page 107-111; 146pp; French.
 XX The papA gene of S.pristinaespiralis is involved in the biosynthesis of 4
 CC -dimethylamino-L-phenylalanine (DMAPA), a precursor for pristinamycin
 CC 1A. Upstream of the papA gene, on the complementary strand, is the snbA
 CC gene coding for 3-hydroxypicolinic acid-AMP ligase. The region between
 CC these two genes was sequenced and two open reading frames were
 CC identified. The first (pipA) decodes to an amino acid sequence with
 CC homology to ornithine cyclodeaminase from Agrobacterium tumefaciens. The
 CC pipA gene product is likely to catalyse the cyclo- deamination of lysine,
 CC leading to production of pipercolic acid. Mutations in the pipA gene were
 CC shown to affect pipercolic acid synthesis but not the synthesis of 3-
 CC hydroxypicolinic acid. The second open reading frame (snbF) could be
 CC decoded to give a product with homology to hydroxylases of the cytochrome
 CC P450 type. Disruption of the pipA and snbF genes can be used to produce
 CC strains of S.pristinaespiralis which are unable to produce the antibiotic
 CC pristinamycin I but which may be able to produce new, modified forms of
 CC it
 XX SQ Sequence 4496 BP; 659 A; 1980 C; 1321 G; 536 T; 0 U; 0 Other;
 Query Match 90.6%; Score 15.4; DB 2; Length 4496;
 Best Local Similarity 94.1%; Pred. No. 9.2e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 TCCGCCGGGTCCGGGAC 17
 Db 2299 TCCGCCGGGTCGCGGAC 2283

Search completed: November 7, 2005, 05:40:12
 Job time : 208.043 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 19:19:37 ; Search time 57.6915 Seconds
(without alignment)
482.163 Million cell updates/sec

Title: US-10-777-131A-2

Perfect score: 17

Sequence: 1 tccgcgggtcggggac 17

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*

4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*

5: /cgn2_6/ptodata/1/ina/PTCUB_COMB.seq.*

6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17	100.0	4403765	3 US-09-103-840A-2	Sequence 2, Appli
C 2	17	100.0	4411529	3 US-09-103-840A-1	Sequence 1, Appli
C 3	15.4	90.6	1194	3 US-08-765-907A-9	Sequence 9, Appli
C 4	15.4	90.6	1194	4 US-09-987-614A-9	Sequence 9, Appli
C 5	15.4	90.6	4496	3 US-08-765-907A-6	Sequence 6, Appli
C 6	15.4	90.6	4496	4 US-09-987-614A-6	Sequence 6, Appli
C 7	15.4	90.6	54986	4 US-09-949-016-16716	Sequence 16716, A
C 8	15	88.2	1167	4 US-09-266-965-97	Sequence 97, Appli
C 9	15	88.2	2969	4 US-09-799-451-388	Sequence 388, App
C 10	15	88.2	3765	4 US-09-266-965-9	Sequence 9, Appli
C 11	15	88.2	53500	4 US-09-266-965-76	Sequence 76, Appli
C 12	15	88.2	4403765	3 US-09-103-840A-2	Sequence 2, Appli
C 13	15	88.2	4411529	3 US-09-103-840A-1	Sequence 1, Appli
C 14	14.4	84.7	507	4 US-09-902-540-5663	Sequence 5663, Ap
C 15	14.4	84.7	765	4 US-09-252-991A-7684	Sequence 7684, Ap
C 16	14.4	84.7	891	4 US-09-724-797-75	Sequence 75, Appli
C 17	14.4	84.7	1032	4 US-09-252-991A-9565	Sequence 9565, Ap
C 18	14.4	84.7	1075	2 US-08-805-860-1	Sequence 1, Appli
C 19	14.4	84.7	1281	4 US-09-252-991A-7879	Sequence 7879, Ap
C 20	14.4	84.7	1344	4 US-09-252-991A-9345	Sequence 9345, Ap
C 21	14.4	84.7	1614	4 US-09-252-991A-9523	Sequence 9523, Ap
C 22	14.4	84.7	2112	4 US-09-252-991A-9429	Sequence 9429, Ap
C 23	14.4	84.7	2313	4 US-09-252-991A-7945	Sequence 7945, Ap
C 24	14.4	84.7	2454	4 US-09-252-991A-9316	Sequence 9316, Ap
C 25	14.4	84.7	2647	4 US-09-220-132-77	Sequence 77, Appli
C 26	14.4	84.7	2647	5 PCT-US93-06251-77	Sequence 77, Appli
C 27	14.4	84.7	3173	4 US-09-799-451-428	Sequence 428, App

C 28	14.4	84.7	3390	4 US-09-902-540-6647	Sequence 6647, Ap
C 29	14.4	84.7	3393	4 US-09-902-540-514	Sequence 514, App
C 30	14.4	84.7	3737	4 US-09-747-371-1	Sequence 1, Appli
C 31	14.4	84.7	5074	4 US-09-902-540-691	Sequence 691, App
C 32	14.4	84.7	10892	4 US-09-902-540-962	Sequence 962, App
C 33	14.4	84.7	12237	4 US-09-949-016-17312	Sequence 17312, A
C 34	14.4	84.7	38155	2 US-09-453-702B-79	Sequence 79, Appli
C 35	14.4	84.7	43280	3 US-08-804-237C-1	Sequence 1, Appli
C 36	14.4	84.7	48908	3 US-09-453-702B-137	Sequence 137, App
C 37	14.4	84.7	51354	4 US-09-902-540-1270	Sequence 1270, App
C 38	14.4	84.7	61158	4 US-09-949-016-15041	Sequence 15041, A
C 39	14.4	84.7	75431	4 US-09-949-016-15122	Sequence 15122, A
C 40	14	82.4	340	4 US-09-513-999C-20103	Sequence 20103, A
C 41	14	82.4	465	3 US-09-188-930-49	Sequence 49, Appli
C 42	14	82.4	465	4 US-09-312-283C-49	Sequence 49, Appli
C 43	14	82.4	601	4 US-09-949-016-67201	Sequence 67201, A
C 44	14	82.4	601	4 US-09-949-016-67202	Sequence 67202, A
C 45	14	82.4	833	3 US-09-188-930-247	Sequence 247, App

ALIGNMENTS

```
RESULT 1
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      100.0%; Score 17; DB 3; Length 4403765;
Best local similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      1 TCCGCCGGTCCGGGAC 17
Db      1286681 TCCGCCGGTCCGGGAC 1286665

RESULT 2
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match          100.0%; Score 17; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TCCGCCGGGTCCGGGAC 17
Db      1287212 TCCGCCGGGTCCGGGAC 1287196

RESULT 3
US-08-765-907A-9/c
; Sequence 9, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765,907A
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-9

Query Match          90.6%; Score 15.4; DB 3; Length 1194;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCCGCCGGGTCCGGGAC 17
Db      500 TCCGCCGGGTCCGGGAC 484

RESULT 4
US-09-987-614A-9/c
; Sequence 9, Application US/09987614A
; Patent No. 6833382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/09/987,614A
```

```

; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US/08/765,907
; PRIOR FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-9

Query Match          90.6%; Score 15.4; DB 4; Length 1194;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCCGCCGGGTCCGGGAC 17
Db      500 TCCGCCGGGTCCGGGAC 484

RESULT 5
US-08-765-907A-6/c
; Sequence 6, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765,907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4496
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-6

Query Match          90.6%; Score 15.4; DB 3; Length 4496;
Best Local Similarity 94.1%; Pred. No. 2.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCCGCCGGGTCCGGGAC 17
Db      2299 TCCGCCGGGTCCGGGAC 2283

RESULT 6
US-09-987-614A-6/c
; Sequence 6, Application US/09987614A
; Patent No. 6833382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
```

;; TITLE OF INVENTION: Mutaenthesis
;; FILE REFERENCE: Streptogramin genes
;; CURRENT APPLICATION NUMBER: US/09/987,614A
;; PRIOR FILING DATE: 2001-11-15
;; PRIOR APPLICATION NUMBER: US/08/765,907
;; PRIOR FILING DATE: 1997-03-20
;; NUMBER OF SEQ ID NOS: 17
;; SOFTWARE: Patent Ver. 2.0
;; SEQ ID NO 6
;; LENGTH: 4496
;; TYPE: DNA
;; ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-6

Query Match 90.6%; Score 15.4; DB 4; Length 4496;
Best Local Similarity 94.1%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCGCGGGTCCGGGAC 17
Db 2299 TCCGCGGGTCCGGGAC 2283

RESULT 7
US-09-949-016-16716/C
;; Sequence 16716, Application US/09949016
;; Patent No. 6812339
;; GENERAL INFORMATION:
;; APPLICANT: VENTER, J. Craig et al.
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;; FILE REFERENCE: CLO01307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; CURRENT FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 16716
;; LENGTH: 54986
;; TYPE: DNA
;; ORGANISM: Human
;; FEATURE:
;; NAME/KEY: misc.feature
;; LOCATION: (1)...(54986)
;; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16716

Query Match 90.6%; Score 15.4; DB 4; Length 54986;
Best Local Similarity 94.1%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCGCGGGTCCGGGAC 17
Db 2310 TCCGCGGGTCCGGGAC 2294

RESULT 8
US-09-266-965-97
;; Sequence 97, Application US/09266965
;; Patent No. 6495348
;; GENERAL INFORMATION:
;; APPLICANT: Sherman, D
;; APPLICANT: Mao, Y
;; APPLICANT: Varoglu, M
;; APPLICANT: He, M
;; APPLICANT: Sheldon, P
;; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
;; FILE REFERENCE: 600.456US1

;; CURRENT APPLICATION NUMBER: US/09/266,965
;; CURRENT FILING DATE: 1999-03-12
;; EARLIER APPLICATION NUMBER: US 08/624,447
;; EARLIER FILING DATE: 1996-08-19
;; EARLIER APPLICATION NUMBER: PCT/US94/11279
;; EARLIER FILING DATE: 1994-10-06
;; EARLIER APPLICATION NUMBER: US 08/133,963
;; EARLIER FILING DATE: 1993-10-07
;; NUMBER OF SEQ ID NOS: 145
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 97
;; LENGTH: 1167
;; TYPE: DNA
;; ORGANISM: Streptomyces lavendulae
US-09-266-965-97

Query Match 88.2%; Score 15; DB 4; Length 1167;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCGCGGGTCCGGGA 16
Db 519 CCGCGGGTCCGGGA 533

RESULT 9
US-09-799-451-388/C
;; Sequence 388, Application US/09799451
;; Patent No. 6783969
;; GENERAL INFORMATION:
;; APPLICANT: Tang, Y. Tom
;; APPLICANT: Zhou, Ping
;; APPLICANT: Goodrich, Ryle
;; APPLICANT: Asundi, Vinod
;; APPLICANT: Ren, Feiyan
;; APPLICANT: Zhang, Jie
;; APPLICANT: Xue, Aidong J.
;; APPLICANT: Zhao, Qing A.
;; APPLICANT: Wang, Jian-Rui
;; APPLICANT: Ma, Yungqing
;; APPLICANT: Yamazaki, Victoria
;; APPLICANT: Chen, Rui-hong
;; APPLICANT: Wang, Zhiwei
;; APPLICANT: Wang, Dunrui
;; APPLICANT: Yang, Yonghong
;; APPLICANT: Wentman, Tom
;; APPLICANT: Ghosh, Reena
;; APPLICANT: Drmanac, Radoje T.
;; TITLE OF INVENTION: No. 6783969e1 Nucleic Acids and
;; FILE REFERENCE: Polypeptides
;; CURRENT APPLICATION NUMBER: US/09/799,451
;; CURRENT FILING DATE: 2001-03-05
;; NUMBER OF SEQ ID NOS: 948
;; SOFTWARE: pc FL_genes Version 2.0
;; SEQ ID NO 388
;; LENGTH: 29659
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (567)..(2711)
US-09-799-451-388

Query Match 88.2%; Score 15; DB 4; Length 2969;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCGCGGGTCCGGGA 16
Db 2818 CCGCGGGTCCGGGA 2804

RESULT 10
US-09-266-965-9
Sequence 9, Application US/09266965
Patent No. 6495348
GENERAL INFORMATION:
APPLICANT: Sherman, D
APPLICANT: Mao, Y
APPLICANT: Varoglu, M
APPLICANT: He, M
APPLICANT: Sheldon, P
TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
FILE REFERENCE: 600.456US1
CURRENT APPLICATION NUMBER: US/09/266.965
CURRENT FILING DATE: 1999-03-12
EARLIER APPLICATION NUMBER: US 08/624,447
EARLIER FILING DATE: 1996-08-19
EARLIER APPLICATION NUMBER: PCT/US94/11279
EARLIER FILING DATE: 1994-10-06
EARLIER APPLICATION NUMBER: US 08/133,963
EARLIER FILING DATE: 1993-10-07
NUMBER OF SEQ ID NOS: 145
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 9
LENGTH: 3765
TYPE: DNA
ORGANISM: Streptomyces lavendulae
US-09-266-965-9

Query Match 88.2%; Score 15; DB 4; Length 3765;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CcGCCGGGTCCGGGGA 16
Db 1097 CCGCGGGTCCGGGGA 1111

RESULT 11
US-09-266-965-76/c
Sequence 76, Application US/09266965
Patent No. 6495348
GENERAL INFORMATION:
APPLICANT: Sherman, D
APPLICANT: Mao, Y
APPLICANT: Varoglu, M
APPLICANT: He, M
APPLICANT: Sheldon, P
TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
FILE REFERENCE: 600.456US1
CURRENT APPLICATION NUMBER: US/09/266.965
CURRENT FILING DATE: 1999-03-12
EARLIER APPLICATION NUMBER: US 08/624,447
EARLIER FILING DATE: 1996-08-19
EARLIER APPLICATION NUMBER: PCT/US94/11279
EARLIER FILING DATE: 1994-10-06
EARLIER APPLICATION NUMBER: US 08/133,963
EARLIER FILING DATE: 1993-10-07
NUMBER OF SEQ ID NOS: 145
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 76
LENGTH: 53500
TYPE: DNA
ORGANISM: Streptomyces lavendulae
US-09-266-965-76

Query Match 88.2%; Score 15; DB 4; Length 53500;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CcGCCGGGTCCGGGGA 16
Db 20939 CCGCGGGTCCGGGGA 20925

RESULT 12
US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103.840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 88.2%; Score 15; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CcGCCGGGTCCGGGAC 17
Db 1306570 CCGCGGGTCCGGGAC 1306584

RESULT 13
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103.840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 88.2%; Score 15; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query 3 CCGCGGGTCCGGGAC 17
Db 1307101 CCGCGGGTCCGGGAC 1307115
RESULT 14
US-09-902-540-5663
Sequence 5663, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:


```
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO: 5663
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-5663

Query Match      84.7%; Score 14.4; DB 4; Length 507;
Best Local Similarity 93.8%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 TCCGCCGGGTCCGGGA 16
        |||||
Db      293 TCCGCCGGGTCCGGGA 308

RESULT 15
US-09-252-991A-7684/C
; Sequence 7684, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 7684
; LENGTH: 765
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7684

Query Match      84.7%; Score 14.4; DB 4; Length 765;
Best Local Similarity 93.8%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 CCGCCGGGTCCGGGAC 17
        |||||
Db      712 CCGCCGGGTCCGGTAC 697
```

Search completed: November 6, 2005, 23:10:09
Job time : 92.6915 secs

This Page Blank (copying)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 20:56:42 ; Search time 423.915 Seconds
(without alignments)
331.640 Million cell updates/sec

Title: US-10-777-131A-2

Perfect score: 17
Sequence: 1 tccgcgggtcggggac 17

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 413490567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
23: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
24: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
25: /cgn2_6/ptodata/2/pubpna/US11_PUBCOMB.seq.*
26: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
27: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
28: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	100.0	17	US-10-216-817-2	Sequence 2, Appli
2	17	100.0	17	US-10-777-131A-2	Sequence 2, Appli
3	17	100.0	823	US-10-216-817-19	Sequence 19, Appli
4	17	100.0	823	US-10-777-131A-19	Sequence 19, Appli
5	15.4	90.6	504	US-10-425-115-163326	Sequence 163326,

C 6	15.4	90.6	879	18	US-10-282-122A-11189	Sequence 11189, A
C 7	15.4	90.6	1368	18	US-10-282-122A-13252	Sequence 13252, A
C 8	15.4	90.6	1415	24	US-10-631-467-1201	Sequence 1201, Ap
C 9	15.4	90.6	1758	16	US-10-156-761-6524	Sequence 6524, Ap
C 10	15.4	90.6	9025608	16	US-10-156-761-1	Sequence 1, Appli
C 11	15.4	90.6	9025608	16	US-10-156-761-1	Sequence 1, Appli
C 12	15	88.2	442	21	US-10-425-115-130735	Sequence 130735,
C 13	15	88.2	474	14	US-10-027-632-195891	Sequence 195891,
C 14	15	88.2	474	18	US-10-027-632-195891	Sequence 195891,
C 15	15	88.2	1062	19	US-10-424-599-7197	Sequence 7197, A
C 16	15	88.2	1167	16	US-10-267-255-97	Sequence 97, Appl
C 17	15	88.2	1167	16	US-10-267-255-97	Sequence 97, Appl
C 18	15	88.2	1518	16	US-10-156-761-5098	Sequence 5098, Ap
C 19	15	88.2	1524	20	US-10-437-963-9032	Sequence 9032, Ap
C 20	15	88.2	1944	20	US-10-437-963-9032	Sequence 9032, Ap
C 21	15	88.2	2969	19	US-10-302-172-388	Sequence 388, App
C 22	15	88.2	3765	16	US-10-267-255-9	Sequence 9, Appli
C 23	15	88.2	3765	16	US-10-267-255-9	Sequence 9, Appli
C 24	15	88.2	53500	10	US-09-953-348-76	Sequence 76, Appl
C 25	15	88.2	53500	16	US-10-267-255-76	Sequence 76, Appl
C 26	14.4	84.7	190	21	US-10-425-115-41023	Sequence 41023, A
C 27	14.4	84.7	250	9	US-09-764-877-2893	Sequence 2893, Ap
C 28	14.4	84.7	250	18	US-10-242-515-2893	Sequence 2893, Ap
C 29	14.4	84.7	310	19	US-10-424-599-76802	Sequence 76802, A
C 30	14.4	84.7	412	20	US-10-437-963-87400	Sequence 87400, A
C 31	14.4	84.7	417	21	US-10-425-115-122422	Sequence 122422,
C 32	14.4	84.7	460	21	US-10-425-115-89259	Sequence 89259, A
C 33	14.4	84.7	464	10	US-09-918-995-8853	Sequence 8853, Ap
C 34	14.4	84.7	509	21	US-10-425-115-19357	Sequence 19357, A
C 35	14.4	84.7	524	19	US-10-424-599-102950	Sequence 102950,
C 36	14.4	84.7	554	20	US-10-767-701-28617	Sequence 28617, A
C 37	14.4	84.7	561	20	US-10-437-963-87275	Sequence 87275, A
C 38	14.4	84.7	563	20	US-10-767-701-21299	Sequence 21299, A
C 39	14.4	84.7	574	19	US-10-424-599-21354	Sequence 21354, A
C 40	14.4	84.7	606	16	US-10-156-761-1573	Sequence 1573, Ap
C 41	14.4	84.7	606	16	US-10-156-761-5910	Sequence 5910, Ap
C 42	14.4	84.7	612	21	US-10-425-115-56398	Sequence 56398, A
C 43	14.4	84.7	633	14	US-10-027-632-285049	Sequence 285049,
C 44	14.4	84.7	633	18	US-10-027-632-285049	Sequence 285049,
C 45	14.4	84.7	636	22	US-10-487-901-6727	Sequence 6727, Ap

ALIGNMENTS

RESULT 1
US-10-216-817-2
; Sequence 2, Application US/10216817
; Publication No. US20030129619A1
GENERAL INFORMATION:
APPLICANT: GICQUEL, BRIGITTE
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
FILE REFERENCE: 03495, 0233-00000
CURRENT APPLICATION NUMBER: US/10-216, 817
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: 60/311, 824
PRIOR FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: 60/313, 523
PRIOR FILING DATE: 2001-08-21
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 17
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Description of Artificial Sequence: primer
US-10-216-817-2
Query Match
Best Local Similarity 100.0%; Score 17; DB 16; Length 17;
Pred. No. 3.1e+02;


```
; Sequence 11189, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11189
; LENGTH: 879
; TYPE: DNA
; ORGANISM: Burkholderia cepacia
US-10-282-122A-11189

Query Match          90.6%; Score 15.4; DB 18; Length 879;
Best Local Similarity 94.1%; Pred. No. 4.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 1 TCCGCCGGGTGCGGAC 17
    |||||
Db 53 TCCGCCGGGTGCGGAC 37

RESULT 7
US-10-282-122A-13252/c
; Sequence 13252, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
```

```
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13252
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Burkholderia fungorum
US-10-282-122A-13252

Query Match          90.6%; Score 15.4; DB 18; Length 1368;
Best Local Similarity 94.1%; Pred. No. 3.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 1 TCCGCCGGGTGCGGAC 17
    |||||
Db 467 TCCGCCGGGTGCGGAC 451

RESULT 8
US-10-631-467-1201
; Sequence 1201, Application US/10631467
; Publication No. US20050208496A1
; GENERAL INFORMATION:
; APPLICANT: Genex Research Inc.
; TITLE OF INVENTION: Method for testing for bronchreal asthma, or chronic obstructive p
; FILE REFERENCE: 3462.1005-000
; CURRENT APPLICATION NUMBER: US/10/631,467
; PRIOR FILING DATE: 2003-07-31
; PRIOR APPLICATION NUMBER: JP 2003-077212
; PRIOR FILING DATE: 2003-03-20
; PRIOR APPLICATION NUMBER: JP 2002-229312
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 2086
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1201
; LENGTH: 1415
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-631-467-1201

Query Match          90.6%; Score 15.4; DB 24; Length 1415;
Best Local Similarity 94.1%; Pred. No. 3.6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 1 TCCGCCGGGTGCGGAC 17
    |||||
Db 73 TCCGCCGGGTGCGGAC 89
```

RESULT 9

US-10-156-761-6524/c
; Sequence 6524, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT FILING DATE: 2002-05-29
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6524
; LENGTH: 1758
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1758)
US-10-156-761-6524

Query Match Best Local Similarity 90.6%; Score 15.4; DB 16; Length 1758;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCCGCCGGGTCCGGGAC 17
DB 1055 TCCGCCGGGTCCGGGAC 1039

RESULT 10

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT FILING DATE: 2002-05-29
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 90.6%; Score 15.4; DB 16; Length 9025608;

Best Local Similarity 94.1%; Pred. No. 17;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCCGCCGGGTCCGGGAC 17
DB 1815959 TCCGCCGGGTCCGGGAC 1815975

RESULT 11

US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT FILING DATE: 2002-05-29
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match Best Local Similarity 90.6%; Score 15.4; DB 16; Length 9025608;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCCGCCGGGTCCGGGAC 17
DB 7838294 TCCGCCGGGTCCGGGAC 7838278

RESULT 12

US-10-425-115-130735/c
; Sequence 130735, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: LA ROSE, THOMAS J.
; APPLICANT: KOVALLIC, DAVID K.
; APPLICANT: ZHOU, YIHUA
; APPLICANT: CAO, YONGWEI
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 130735
; LENGTH: 442
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(442)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MWT4577_50707C.1

Query Match 90.6%; Score 15.4; DB 16; Length 9025608;

US-10-425-115-130735

Query Match 88.2% Score 15; DB 21; Length 442;

Best Local Similarity 100.0%; Pred. No. 8.3e+02; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGCCGGGTCCGGG 15
|||||
Db 385 TCCGCCGGGTCCGGG 371

RESULT 13

US-10-027-632-195991/c

; Sequence 195991, Application US/10027632

; Publication No. US20020198371A1

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO: 195991

; LENGTH: 474

; TYPE: DNA

; ORGANISM: Human

; US-10-027-632-195991

Query Match 88.2% Score 15; DB 14; Length 474;

Best Local Similarity 100.0%; Pred. No. 8.1e+02; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CCGCCGGGTCCGGGA 16
|||||
Db 161 CCGCCGGGTCCGGGA 147

RESULT 14

US-10-027-632-195991/c

; Sequence 195991, Application US/10027632

; Publication No. US20030204075A9

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO: 195991

; LENGTH: 474

; TYPE: DNA

; ORGANISM: Human

; US-10-027-632-195991

Query Match 88.2% Score 15; DB 18; Length 474;

Best Local Similarity 100.0%; Pred. No. 8.1e+02; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CCGCCGGGTCCGGGA 16
|||||
Db 161 CCGCCGGGTCCGGGA 147

RESULT 15

US-10-424-599-27197

; Sequence 27197, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO: 27197

; LENGTH: 1002

; TYPE: DNA

; ORGANISM: Glycine max

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)..(1002)

; OTHER INFORMATION: unsure at all n locations

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_12455C.1

; US-10-424-599-27197

Query Match 88.2% Score 15; DB 19; Length 1002;

Best Local Similarity 100.0%; Pred. No. 6.3e+02; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCGCCGGGTCCGGG 15
|||||
Db 77 TCCGCCGGGTCCGGG 91Search completed: November 7, 2005, 05:21:08
Job time : 459.915 secs

It's Page Blank (uspo)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 02:38:52 ; Search time 1700.72 Seconds
(without alignments)
380.481 Million cell updates/sec

Title: US-10-777-131A-2

Perfect score: 17

Sequence: 1 tccgcgcgggtccggggac 17

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 6847908

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_est1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	94.1	366	5	BY111757 BY111757
2	16	94.1	438	2	BT776822 MY-20-H-0
3	16	94.1	720	4	BG857100 1024049H0
4	16	94.1	932	5	B0838754 AGENCOURT
5	15.4	90.6	193	1	AI316193 u126b02.Y
6	15.4	90.6	217	5	BY000112 BY000112
7	15.4	90.6	218	2	BB561944 BB561944
8	15.4	90.6	218	5	BY000049 BY000049
9	15.4	90.6	219	9	AT418508 Mus muscu
10	15.4	90.6	222	5	BY006303 BY006303
11	15.4	90.6	228	5	BY006302 BY006302
12	15.4	90.6	234	2	BB561733 BB561733
13	15.4	90.6	236	2	BB561740 BB561740
14	15.4	90.6	243	1	AU096437 AU096437
15	15.4	90.6	244	1	AA277884 VC10C05.1
16	15.4	90.6	247	5	BY000013 BY000013
17	15.4	90.6	248	2	BB564084 BB564084
18	15.4	90.6	250	5	BY339805 BY339805
19	15.4	90.6	251	5	BY338153 BY338153
20	15.4	90.6	254	5	BY006482 BY006482
21	15.4	90.6	265	2	BB562086 BB562086
22	15.4	90.6	265	2	BB564146 BB564146
23	15.4	90.6	266	2	BB561699 BB561699
24	15.4	90.6	268	2	BB561859 BB561859

25	15.4	90.6	270	2	BB561705 BB561705
26	15.4	90.6	271	2	BB561925 BB561925
27	15.4	90.6	272	2	BB561694 BB561694
28	15.4	90.6	274	2	BB561741 BB561741
29	15.4	90.6	274	2	BB561744 BB561744
30	15.4	90.6	274	2	BB561750 BB561750
31	15.4	90.6	274	2	BB561845 BB561845
32	15.4	90.6	277	2	BB561730 BB561730
33	15.4	90.6	277	2	BB562206 BB562206
34	15.4	90.6	277	2	BB562239 BB562239
35	15.4	90.6	281	5	BY318477 BY318477
36	15.4	90.6	283	5	BY034687 BY034687
37	15.4	90.6	292	2	BB562015 BB562015
38	15.4	90.6	295	5	BY336579 BY336579
39	15.4	90.6	309	6	BY789788 BY789788
40	15.4	90.6	313	8	BY333785 BY333785
41	15.4	90.6	313	8	CC200248 CC200248
42	15.4	90.6	316	5	BY000114 BY000114
43	15.4	90.6	324	6	CD552437 CD552437
44	15.4	90.6	325	5	BY151676 BY151676
45	15.4	90.6	326	5	BY346975 BY346975

ALIGNMENTS

RESULT 1

LOCUS

BY111757 RIKEN full-length enriched, stroma cell Mus musculus cDNA

clone I320023J05 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

COMMENT

TITLE

JOURNAL

MEDLINE

PubMed

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan.
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/

Altawane, K., Akimura, T., Aikawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN Integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Takashi Ishikawa (Department of Surgery
2 Yokohama City University 3-9 Fukuura, Kanazawa-ku, Yokohama
236-0004 Japan) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

FEATURES

source

location/Qualifiers
1..366
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="R320023J05"
/cell_type="stroma cell"
/clone_lib="RIKEN full-length enriched, stroma cell"

ORIGIN

Query Match 94.1%; Score 16; DB 5; Length 366;
Best Local Similarity 94.1%; Pred. NO. 3.2e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 1 TCCGCCGGGTCGGGAC 17
|||||
Db 98 TCCGCCGGGTCGGGAC 114

RESULT 2
BE776822 438 bp mRNA linear EST 20-SEP-2000
LOCUS MY-20-H-01 Plinfestansmy Phytophthora infestans cDNA, mRNA sequence.
ACCESSION BE776822
VERSION BE776822.1 GI:10230477
KEYWORDS EST.
SOURCE Phytophthora infestans (potato late blight agent)
ORGANISM Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.

REFERENCE
1 (bases 1 to 438)
Kamoun, S., Hraber, P. T., Sobral, B. W. S., Nuss, D. and Govers, F.
Initial assessment of gene diversity for the oomycete pathogen
Phytophthora infestans based on expressed sequences
Fungal Genet. Biol. 28 (2), 94-106 (1999)
20056376
JOURNAL MEDLINE
10587472
PUBMED
Contact: Govers F
Laboratory of Phytopathology

Wageningen University
Binnenhaven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands
Tel: 31 317 483 138
Fax: 31 317 483 412
Email: Francine.Govers@medew.fyco.wau.nl.
Location/Qualifiers
1..438
/organism="Phytophthora infestans"
/mol_type="mRNA"
/strain="DDR7602, A1 mating type"
/db_xref="taxon:4787"
/dev_stage="4-week old vegetative, non-sporulating
mycelium in synthetic medium"
/lab_host="E. coli, strain DHS-alpha"
/clone_lib="Plinfestansmy"
/note=Vector: pSPOR1; Site 1: SalI; Site 2: NotI; Total
RNA was isolated from mycelium of P. infestans DDR7602
cultured for 4 weeks in synthetic medium. EST clones were
named by their position in the microtiter plate, preceded
by the prefix MY (for mycelial) and the successive number
of the microtiter plate (e.g. MY-06-A-04)."

FEATURES

source

Query Match 94.1%; Score 16; DB 2; Length 438;
Best Local Similarity 94.1%; Pred. NO. 3.1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 1 TCCGCCGGGTCGGGAC 17
|||||
Db 233 TCCGCCGGGTCGGGAC 249

ORIGIN

RESULT 3
BG857100 720 bp mRNA linear EST 29-MAY-2001
LOCUS 1024049H09.x1 C. reinhardtii CC-1690, normalized, lambda Zap II
DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BG857100
VERSION BG857100.1 GI:14238284
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonadae.

REFERENCE
1 (bases 1 to 720)
Grossman, A., Davies, J., Federpiel, N., Harris, B., Lefebvre, P.,
McDermott, J. P., Sillflow, C., Stern, D. and Strzyski, R.
Analyses of the Chlamydomonas reinhardtii genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 2
Unpublished (2000)
Contact: Charles Hauser
Duke University
Duke Box 91000
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauuser@duke.edu.

FEATURES

source

Location/Qualifiers
1..720
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, lambda Zap
II"
/note=Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light, TAP medium in the dark, HS (minimal) medium in
ambient levels of CO2 and HS medium bubbled with 5% CO2.
PolyA mRNA was purified from each sample, pooled and cDNA

synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

ORIGIN

Query Match 94.1%; Score 16; DB 4; Length 720;
Best Local Similarity 100.0%; Pred. No. 2.9e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCGCGGGTCCGGGAG 17
|||||
DB 479 CCGCGGGTCCGGGAG 494

RESULT 4
LOCUS BU838754/C 932 bp mRNA linear EST 16-OCT-2002
DEFINITION AGENCOURT 8210078 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6258013
5' mRNA sequence.
ACCESSION BU838754
VERSION BU838754.1 GI:24023149
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (baaes 1 to 932)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CNA Library Preparation: Rubin Laboratory
CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2413 row: d column: 14
High quality sequence stop: 543.
Location/Qualifiers

FEATURES
SOURCE 1..932
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6258013"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_112"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 94.1%; Score 16; DB 5; Length 932;
Best Local Similarity 100.0%; Pred. No. 2.8e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCGCGGGTCCGGGAG 17
|||||
DB 930 CCGCGGGTCCGGGAG 915

RESULT 5

A1316193
LOCUS A1316193 193 bp mRNA linear EST 17-DEC-1998
DEFINITION u12eb02.y1 Sugano mouse kidney m1a Mus musculus cDNA clone IMAGE:1921035 5' similar to gb:X58295_rnai PLASMA GLUTATHIONE PEROXIDASE PRECURSOR (HUMAN) ; gb:U13705 Mus musculus domesticus C57BL/6J plasma glutathione (mouse); mRNA sequence.

ACCESSION A1316193
VERSION A1316193.1 GI:4031460
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
MUS musculus

REFERENCE 1 (bases 1 to 193)
AUTHORS Marmè,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:97327
Seq primer: custom primer used
High quality sequence stop: 192.
Location/Qualifiers

FEATURES
SOURCE 1..193
Location/Qualifiers

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1921035"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Sugano mouse kidney m1a"
/note="Organ: kidney; Vector: pMB18S-FL3; Site 1: DraIII (CACTGTG); Site 2: DraIII (CACTGTG); 1st strand cDNA was primed with an oligo(dT) primer (ATGTGGCTTTTCTTTTCTTTT) ; double-stranded cDNA was ligated to a DraIII adaptor (TTGTGGCTTACTGG) , digested and cloned into distinct DraIII sites of the pMB18S-FL3 vector (5' site CACTGTG, 3' site CACTGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCGCTCTTAACCTGG and 3' end primer CGACCTGACCTGACGACA."

ORIGIN

Query Match 90.6%; Score 15.4; DB 1; Length 193;
Best Local Similarity 94.1%; Pred. No. 6.7e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCGCGGGTCCGGGAG 17
|||||
DB 96 TCCGCGGGTCCGGGAG 112

RESULT 6
LOCUS BY000112 217 bp mRNA linear EST 06-DEC-2002
DEFINITION BY000112 RIKEN full-length enriched, adult male kidney Mus musculus cDNA clone 0610007p21 5', mRNA sequence.

ACCESSION	BY000112
VERSION	BY000112.1
KEYWORDS	GI:26060361
SOURCE	EST.
ORGANISM	Mus musculus (house mouse)
REFERENCE	Mus musculus
AUTHORS	Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 217)
	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikiado, I., Oono, N., Saito, R., Suzuki, H., Yamataka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bull, C. C., Hume, D. A., Quackenbush, J., Schmitt, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chochia, C., Corbani, L. E., Cousins, S., Dalla, E., Dargatz, T. A., Fletcher, C. F., Forrest, A., Fraser, K. S., Gaasterland, T., Gariboldi, M., Glessi, C., Godzik, A., Gough, J., Grimmond, S., Guenichon, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., Kling, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lehar, B., Lyons, P. A., Maglott, D. R., Maltale, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numa, K., Okido, T., Pavan, W. J., Perle, G., Pezle, G., Petrovsky, N., Pillai, R., Pontus, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringald, M., Sadelin, A., Schneider, C., Semp, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wymshew-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavalnik, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Sakazume, T., Fukuda, S., Hara, A., Heshizume, W., Imotani, K., Ishii, Y., Aikawa, I., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEIDLINE	22354683
PUBMED	12466851
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suenho-cho, Tsunumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-ree@sec.riken.jp, URL: http://genome.gsc.riken.jp/ Aizawa, K., Akimura, T., Aikawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Nunazaki, R., Ono, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome, 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1571-1577 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site (http://genome.gsc.riken.go.jp) for further details. Location/Qualifiers

source	1. 217	/organism="Mus musculus"
		/mol_type="mrna"
		/db_xref="taxon:10090"
		/clone="0610007p21"
		/sex="male"
		/tissue_type="kidney"
		/dev_stage="adult"
		/lab_host="SOLR"
		/note="Site 1: XhoI; Site 2: SacI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5']
		GAGAGAGAGCGCGCGCGACACTGAGTTTATTTTATTTTATN 3'}. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5']
		GAGAGAGAGAGAGTCCAGAGCGTCATTTATTTATTAACCCCCCCCC 3'}. cDNA was cleaved with XhoI and SacI."
ORIGIN		
Query Match	90.6%;	Score 15.4; DB 5; Length 217;
Best Local Similarity	94.1%;	Pred. No. 6.6e+03;
Matches	16; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
Oy	1	TCGCGCGCGTCGGGGAC 17
Db	161	TCGCGCGCGCGCGGGAC 177
RESULT 7		
LOCUS	BB561944	218 bp mRNA Linear EST 29-NOV-2000
DEFINITION	BB561944 RIKEN full-length enriched, adult male kidney Mus musculus	
ACCESSION	BB561944	cDNA clone 0610005A12 5', mRNA sequence.
VERSION	BB561944.1	GI:11452836
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 218)	
AUTHORS	Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P., Haneagaki,T., Hayatsu,N., Hirooka,T., Hirozane,T., Hodojima,Y., Imotani,K., Ishi,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Komori,K., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toyu,T., Watabiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.	
TITLE	RIKEN Mouse ESTs (Aizawa,K. et al. 2000)	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 220-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.jp, URL:http://genome.gsc.riken.jp/ Sasaki,N., Nishiyama,Y., Westover,A., Itoh,M., Nagao,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. P roc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)	

Itoh, M., Kitanai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers
1. 218
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="0610005A12"
/sex="male"
/tissue_type="kidney"
/dev_stage="adult"
/lab_host="SOLR"
/clone.lib="RIKEN full-length enriched, adult male kidney"
/note="Site 1: XhoI; Site 2: SctI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'.
GAGAGAGAGCGCGCGCACTGAGTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'.
GAGAGAGAGAGATCGAAGCTCAATTATTAATTAACCCCCCCCC 3'.
cDNA was cleaved with XhoI and SctI. "

ORIGIN

Query Match 90.6%; Score 15.4; DB 2; Length 218;
Best Local Similarity 94.1%; Pred. No. 6.6e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCGCCGGGTCGGGAGC 17
|||||
Db 97 TCCGCCGGGTCGGGAGC 113

RESULT 8
BY000049 218 bp mRNA linear EST 06-DEC-2002
LOCUS BY000049 RIKEN full-length enriched, adult male kidney Mus musculus
DEFINITION cDNA clone 0610007122 5', mRNA sequence.
ACCESSION BY000049
VERSION BY000049.1 GI:26060298
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE

1 (bases 1 to 218)
Okazaki, Y., Futuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Buit, C., Hume, D.A., Quackenbush, J., Schiml, L.W., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chochia, C., Corbani, L.E., Cousins, S., Dalla, E., Diegani, T.A., Fletcher, C.F., Forrest, A., Fraser, K.S., Gaasterland, T., Gariboldi, M., Giasi, C., Godzik, A., Gough, J., Grimmond, S., Gutentich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltaise, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,

FEATURES

source

Petrovsky, N., Piljal, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sanelin, A., Schneider, C., Sempke, C.A., Setou, M., Shinada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, L., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kikukawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Yashiro, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.jp URL: <http://genome.gsc.riken.jp/>
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences *Mamm. Genome.* 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers
1. 218
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="0610007122"
/sex="male"
/tissue_type="kidney"
/dev_stage="adult"
/lab_host="SOLR"
/clone.lib="RIKEN full-length enriched, adult male kidney"
/note="Site 1: XhoI; Site 2: SctI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'.
GAGAGAGAGCGCGCGCACTGAGTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'.
GAGAGAGAGAGATCGAAGCTCAATTATTAATTAACCCCCCCCC 3'.
cDNA was cleaved with XhoI and SctI. "

ORIGIN

GAGAGAGAGATCCAGAGCTCATTAATTAATTAACCCCCCCC 3'.
cDNA was cleaved with XhoI and SstI. "

Query Match 90.6%; Score 15.4; DB 5; Length 218;
Best Local Similarity 94.1%; Pred. No. 6.6e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TCCGCGGGGTCCGGGAC 17
Db 98 TCCGCGGGGCGCGGAC 114

RESULT 9
AY18508
LOCUS
DEFINITION
Mus musculus GPX3 gene, VIRTUAL TRANSCRIPT, partial sequence.

ACCESSION
AY18508
VERSION
AY18508.1 GI:39774468
KEYWORDS
GSS.
SOURCE
Mus musculus (house mouse)

REFERENCE
AUTHORS
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tenenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Snihsy, J.J.,
Adams, M.D. and Cargill, M.
Interfering nonneutral evolution from human-chimp-mouse orthologous
gene trios

TITLE
JOURNAL
Science 302 (5652), 1960-1963 (2003)
PUBMED
14671302

REFERENCE
AUTHORS
2 (bases 1 to 219)
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tenenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Snihsy, J.J.,
Adams, M.D. and Cargill, M.
Direct Submission

TITLE
JOURNAL
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA

COMMENT
These sequences were made by sequencing genomic exons and ordering
them based on alignment.

FEATURES
source
1..219
Location/Qualifiers

gene
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>219
/gene="GPX3"
/locus_tag="Hm6580"

ORIGIN

Query Match 90.6%; Score 15.4; DB 9; Length 219;
Best Local Similarity 94.1%; Pred. No. 6.6e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TCCGCGGGGTCCGGGAC 17
Db 57 TCCGCGGGGCGCGGAC 73

RESULT 10
BY006303
LOCUS
DEFINITION
BY006303 RIKEN full-length cDNA clone F520009F08 5', mRNA sequence.
ACCESSION
BY006303
VERSION
BY006303.1 GI:26066552
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 222)
Okazaki, Y., Furuno, M., Kaekawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Otsu, N., Saito, R., Suzuki, H., Yamamoto, T.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Haegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Butt, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Macauld, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Brd, D., Bragan, T.A.,
Chotcha, C., Corbani, L.E., Cousine, S., Delle, E., Dregni, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Guatindich, S., Hirokawa, N., Jackson, I.J., Jarvis, B.D., Kanai, A.,
Kawai, H., Kawasawa, Y., Kedziercki, R.M., King, B.L., Kongsaye, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Nunata, K., Okido, T., Pavan, W.J., Petrea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Warande, Y.,
Wells, C., Wilming, L.G., Wyntaw-Boris, A., Yangisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavoian, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imokani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851

TITLE
JOURNAL
MEDLINE
PUBMED

COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imokani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numasaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

FEATURES
source

1..222
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="F520009F08"

DEFINITION BB561733 RIKEN full-length enriched, adult male kidney Mus musculus
cDNA clone 0600001F01 5', mRNA sequence.

ACCESSION BB561733
VERSION BB561733.1 GI:11452625
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 234)

REFERENCE
AUTHORS Atawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,
Carninci,P., Hanagaki,T., Hayatsu,N., Hirooka,T., Hirozane,T.,
Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,
Kojima,Y., Kono,H., Kusakabe,M., Matsuyama,T., Miyazaki,A.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y.,
Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K.,
Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,
Watabiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoaka,S.,
Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermocactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Tomaru,Y., Carninci,P., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES
source
1..234
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="0600001F01"
/sex="male"
/tissue_type="kidney"
/dev_stage="adult"
/lab_host="SOLR"
/clone_1lb="RIKEN full-length enriched, adult male kidney"
/note="Site 1: XhoI; Site 2: SctI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer 5'
GAGAGAGAGCGCCGCAACGAGTTTCTTTTCTTTTCTTTT 3', cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence 5'
GAGAGAGAGATCCAGAGCATTAATTAATTAACCCCCCCCC 3'.
cDNA was cleaved with XhoI and SctI."

ORIGIN
Query Match 90.6%; Score 15.4; DB 2; Length 234;

Best Local Similarity 94.1%; Pred. No. 6.5e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

cy 1 TCCGCCGGCTCGGGGAC 17
db 98 TCCGCCGGCGCGGGAC 114

RESULT 13
LOCUS BB561740 236 bp mRNA linear EST 29-NOV-2000
DEFINITION BB561740 RIKEN full-length enriched, adult male kidney Mus musculus
cDNA clone 0600001G01 5', mRNA sequence.

ACCESSION BB561740
VERSION BB561740.1 GI:11452632
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 236)

REFERENCE
AUTHORS Atawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,
Carninci,P., Hanagaki,T., Hayatsu,N., Hirooka,T., Hirozane,T.,
Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,
Kojima,Y., Kono,H., Kusakabe,M., Matsuyama,T., Miyazaki,A.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y.,
Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K.,
Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,
Watabiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoaka,S.,
Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermocactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Tomaru,Y., Carninci,P., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES
source
1..236
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="0600001G01"
/sex="male"
/tissue_type="kidney"
/dev_stage="adult"
/lab_host="SOLR"
/clone_1lb="RIKEN full-length enriched, adult male kidney"
/note="Site 1: XhoI; Site 2: SctI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was

This page under review

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:11:16 ; Search time 785.957 Seconds
(without alignments)
1233.024 Million cell updates/sec

Title: US-10-777-131A-3

Perfect score: 20
Sequence: 1 tcgaagtggtgcaaatcgtg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sbs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	6 AX770105	AX770105 Sequence
2	20	100.0	1144	6 AX770122	AX770122 Sequence
3	20	100.0	35336	1 MSGY367	AD000008 Mycobacte
4	20	100.0	103837	1 AEO00516_43	Continuation (44 o
5	20	100.0	244800	1 BX842584	BX842584 Mycobacte
6	20	100.0	278492	1 BX248347	BX248347 Mycobacte
7	18	90.0	349895	1 BX248359	BX248359 Corynebac
8	17.4	87.0	143966	8 AC135920	AC135920 Oryza sat
9	17.4	87.0	204254	10 AC122848	AC122848 Mus muscu
10	17	85.0	176823	9 AC146138	AC146138 Pan trogl
11	17	85.0	178455	2 AC146015	AC146015 Pan trogl
12	16.8	84.0	25454	3 CEY102A5D	AL031628 Caenorhab
13	16.8	84.0	108424	9 AL669827	AL669827 Mouse DNA
14	16.8	84.0	147180	9 AC136469	AC136469 Homo sapi
15	16.8	84.0	156942	2 AC136470	AC136470 Homo sapi
16	16.8	84.0	166618	2 AC010323	AC010323 Homo sapi
17	16.8	84.0	168982	10 AC125324	AC125324 Mus muscu
18	16.8	84.0	189562	2 AC148560	AC148560 Oryctolag
19	16.8	84.0	196855	2 AC022297	AC022297 Mus muscu

C	20	16.8	84.0	202232	2	AC124759	AC124759 Mus muscu
C	21	16.8	84.0	204037	10	AL596207	AL596207 Mouse DNA
C	22	16.8	84.0	211892	10	AL732403	AL732403 Mouse DNA
C	23	16.8	84.0	228645	10	AL450397	AL450397 Mouse DNA
C	24	16.8	84.0	231611	2	AC104870	AC104870 Mus muscu
C	25	16.8	84.0	235041	2	AC130100	AC130100 Rattus no
C	26	16.8	84.0	249984	2	AC139594	AC139594 Rattus no
C	27	16.4	82.0	1890	1	AF352014	AF352014 Vibrio gp
C	28	16.4	82.0	11803	1	AB015554	AB015554 Shewanell
C	29	16.4	82.0	33896	1	AF523728S	AF523728 Azospiril
C	30	16.4	82.0	44226	9	AY130859	AY130859 Homo sapi
C	31	16.4	82.0	44300	2	AC006109	AC006109 Homo sapi
C	32	16.4	82.0	48703	2	AC100448	AC100448 Mus muscu
C	33	16.4	82.0	69399	8	AP006350	AP006350 Lotus cor
C	34	16.4	82.0	87645	8	AP006659	AP006659 Lotus cor
C	35	16.4	82.0	110000	1	AF006841_12	AF006841 Continuation (13 o
C	36	16.4	82.0	111311	9	HS27919	HS27919 Human DNA
C	37	16.4	82.0	125980	9	AC093223	AC093223 Homo sapi
C	38	16.4	82.0	150263	9	AC145132	AC145132 Homo sapi
C	39	16.4	82.0	152464	2	AC027421	AC027421 Homo sapi
C	40	16.4	82.0	157243	9	AL353622	AL353622 Human DNA
C	41	16.4	82.0	157986	2	AC025860	AC025860 Homo sapi
C	42	16.4	82.0	164597	3	AC011907	AC011907 Drosophila
C	43	16.4	82.0	167178	2	AC025672	AC025672 Homo sapi
C	44	16.4	82.0	170575	2	AC019758	AC019758 Drosophila
C	45	16.4	82.0	170977	9	AP000756	AP000756 Homo sapi

ALIGNMENTS

RESULT 1	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
LOCUS	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
DEFINITION	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
ACCESSION	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
VERSION	AX770105.1	GI:32437683	20 bp	DNA	linear	PAT 02-JUL-2003
SOURCE	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
ORGANISM	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
REFERENCE	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
AUTHORS	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
TITLE	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
JOURNAL	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
FEATURES	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
source	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
ORIGIN	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
Query Match	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
Beet Local Similarity	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
Matches	AX770105	Sequence 3 from Patent WO03016562.	20 bp	DNA	linear	PAT 02-JUL-2003
1	TCGAAGTGCGCAATCGTG	20	100.0%	Score 20;	DB 6;	Length 20;
2		20	100.0%	Pred. No. 75;	0;	Indels 0;
3		20	100.0%	Mismatches 0;	0;	Gaps 0;
4		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
5		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
6		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
7		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
8		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
9		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
10		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
11		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
12		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
13		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
14		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
15		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
16		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
17		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
18		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
19		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
20		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
21		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
22		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
23		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
24		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
25		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
26		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
27		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
28		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
29		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
30		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
31		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
32		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
33		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
34		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
35		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
36		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
37		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
38		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
39		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
40		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
41		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
42		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
43		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
44		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
45		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
46		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
47		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
48		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
49		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
50		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
51		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
52		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
53		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
54		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
55		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
56		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
57		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
58		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
59		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
60		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
61		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
62		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
63		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
64		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
65		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
66		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
67		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
68		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
69		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
70		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
71		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
72		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
73		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
74		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
75		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
76		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
77		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
78		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
79		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
80		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
81		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
82		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
83		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
84		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
85		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
86		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
87		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
88		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
89		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
90		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
91		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
92		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
93		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
94		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
95		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
96		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
97		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
98		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
99		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;
100		20	100.0%	Matches 20;	Conservative 0;	Mismatches 0;

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE
1 Glouquet, B.
Compositions and methods for detecting multidrug resistant strains
of M. tuberculosis having mutations in genes of the mutR family
Patent: WO 03016562-A 20 27-FEB-2003;
INSTITUT PASTEUR (FR)

FEATURES
SOURCE
1. 1144
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"

ORIGIN
Query Match 100.0%; Score 20; DB 6; Length 1144;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCGTG 20
7 TCGAAGTGGGCAATCGTG 26

RESULT 3
MSGY367 35336 bp DNA linear BCT 03-DEC-1996
LOCUS Mycobacterium tuberculosis sequence from clone y367.
DEFINITION
ACCESSION AD000008
VERSION AD000008.1 GI:1702971
KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE
1 Du, L.
Direct Submission
Submitted (11-OCT-1996) L. Du, Genome Therapeutics Corporation, 100
Beaver Street, Waltham, MA, USA, 02154 du@ctic.com
COMMENT GSDB:S:1004717.
FEATURES
SOURCE
1. 35336
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"
/clone="y367"

ORIGIN
Query Match 100.0%; Score 20; DB 1; Length 35336;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCGTG 20
14705 TCGAAGTGGGCAATCGTG 14686

RESULT 4
AE000516_43
WPCOMMENT
Sequence split into 44 fragments LOCUS AE000516 Accession AE000516

Fragment Name Begin End
AE000516_00 1 110000
AE000516_01 100001 210000
AE000516_02 200001 310000
AE000516_03 300001 410000
AE000516_04 400001 510000
AE000516_05 500001 610000
AE000516_06 600001 710000
AE000516_07 700001 810000

AE000516_08 800001 910000
AE000516_09 900001 1010000
AE000516_10 1000001 1110000
AE000516_11 1100001 1210000
AE000516_12 1200001 1310000
AE000516_13 1300001 1410000
AE000516_14 1400001 1510000
AE000516_15 1500001 1610000
AE000516_16 1600001 1710000
AE000516_17 1700001 1810000
AE000516_18 1800001 1910000
AE000516_19 1900001 2010000
AE000516_20 2000001 2110000
AE000516_21 2100001 2210000
AE000516_22 2200001 2310000
AE000516_23 2300001 2410000
AE000516_24 2400001 2510000
AE000516_25 2500001 2610000
AE000516_26 2600001 2710000
AE000516_27 2700001 2810000
AE000516_28 2800001 2910000
AE000516_29 2900001 3010000
AE000516_30 3000001 3110000
AE000516_31 3100001 3210000
AE000516_32 3200001 3310000
AE000516_33 3300001 3410000
AE000516_34 3400001 3510000
AE000516_35 3500001 3610000
AE000516_36 3600001 3710000
AE000516_37 3700001 3810000
AE000516_38 3800001 3910000
AE000516_39 3900001 4010000
AE000516_40 4000001 4110000
AE000516_41 4100001 4210000
AE000516_42 4200001 4310000
AE000516_43 4300001 4400000
Continuation (44 of 44) of AE000516 from base 4300001 (AE000516 Mycobacterium tuberculosis)

Query Match 100.0%; Score 20; DB 1; Length 103837;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCGTG 20
85562 TCGAAGTGGGCAATCGTG 85581

RESULT 5
BX842584
LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 13/13.
DEFINITION BX842584 AL021426 AL022076 AL022120 AL123456 Z83864
ACCESSION Z94121 Z97188
VERSION BX842584.1 GI:38490370
KEYWORDS
SOURCE complete genome.
ORGANISM Mycobacterium tuberculosis H37Rv
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE
1 Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Fellwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squares, S., Squares, R., Sultston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.
TITLE Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
JOURNAL Nature 393 (6685), 537-544 (1998)

MEDLINE	98295987
PUBMED	9634230
REFERENCE	2
AUTHORS	Canus,J.C., Pryor,M.J., Medigue,C. and Cole,S.T.
TITLE	Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv
JOURNAL	Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)
MEDLINE	22255591
PUBMED	12368430
REFERENCE	3 (bases 1 to 244800)
AUTHORS	Parkhill,I.J.
TITLE	Direct Submission
JOURNAL	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk On or before Nov 21, 2003 this sequence version replaced gi:3261511, gi:3256026, gi:3261558, gi:3261559, gi:3261648, gi:3261687, gi:3261736, gi:3261805. Notes: Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/). Location/Qualifiers
FEATURES	1..244800
SOURCE	/organism="Mycobacterium tuberculosis H37Rv" /mol_type="genomic DNA" /strain="H37Rv" /db_xref="taxon:83332" complement(89..1396) locus_tag="R3722C" complement(89..1396) locus_tag="R3722c" function="UNKNOWN" note="RV3722c, (MTV025_070c), len: 435 aa. Conserved hypothetical protein, equivalent to O69513 MLCB2407.14 (alias Q9CH73 ML2336..463 aa) HYPOTHETICAL 46.8 KDA PROTEIN from Mycobacterium leprae (426 aa), FASTA scores: opt: 2505, E(): 8.3e-154, (87.25% identity in 424 aa overlap). Also highly similar to Q9RU17 DR1579 CONSERVED HYPOTHETICAL PROTEIN from Deinococcus radiodurans (452 aa), FASTA scores: opt: 1152, E(): 3.1e-67, (44.8% identity in 422 aa overlap); and partially similar to G9I171 P1654 PROBABLE AMINOTRANSFERASE from Pseudomonas aeruginosa (388 aa) FASTA scores: opt: 162, E(): 0.0078, (25.85% identity in 348 aa overlap) and other aminotransferases." Trpase score is 0.900. N-terminus extended since first submission (previously 408 aa)."
CDS	/codon_start=1 /evidence=experimental transl_table=1 product="CONSERVED HYPOTHETICAL PROTEIN" protein_id="CAA18044.2" db_xref="GI:38490371" translation="MSFDSLSQYELAAIHARHOQVAALQGKMLADLTTRGKPSAEQDLDLSQLSLPEDDYRDDEGTSTRNYGGQHGLPGRLAFDELGLAVPLILGNSSLELMHIIVAFSMLYGVSDSPFWIOEDQIKFLCPVDGRHPAIETMGIIEMIPIPMICDGEPVIDVIELVAYDPALIKGMWTVFGNPNPVSVTSRTVRLVIOMRAADPRLPYMLNAVYHLTLTDFPRQVDVLAKAGNPNNPVASPTKTFIPAGGVSFFSGISGNIMWYLOYAGKSKIGPDKNQDLRIIRFRGDADGRLIMLRHQQLIAKFALVAEVDORISRBSRKASWTBPKRGFTISDVLVGTARKRTVALAKOVGLAVTAGASPFRKDPPDKNIRIASFPSPVDLRNVAVDGLATCALLAATETLLNOGLASSAPNVR"
gene	1613..1698 /gene="tRNA-Ser (GGA)" 1613..1698 /gene="tRNA-Ser (GGA)" product="tRNA-Ser" note="codon recognized: UCC" anticodon=(pos:1647..1649,aa:Ser) 1804..2568 locus_tag="R3723"
tRNA	1804..2568 locus_tag="R3723"
CDS	1804..2568

/locus_tag="RV3723"
 /function="UNKNOWN"
 /note="RV3723, (MTV025.071), len: 254 aa. Probable conserved transmembrane protein, with hydrophobic stretches at the N-terminus, and equivalent to 069512|ML2337|MLCB2407.13c PUTATIVE MEMBRANE PROTEIN from Mycobacterium leprae (250 aa), FASTA scores: opt: 1029, E(): 1.2e-44, (64.45% identity in 253 aa overlap). Tbpase score is 0.900."
 /codon_start=1
 /transl_table=11
 /product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN"
 /protein_id="CAE5633.1"
 /db_xref="GI:2860147"
 /translation="MGRKVAALVMAHSPISGAVLYFPVLPPELMGDTGHSIGTGLRLSLADLGLAALPVFTLLTRKRELDTPLDASIMSIWISMAVLGALIVGAIRSEVWLSIDAGQAMFGIYGAAMAAIVAGFPGFSLTSPVAISPPPKPLPKPKORRLRRKATGADGEAPPAEAPEATLEAOEDEAEAPAESESPGSPESATREAPAAETPATAAEERGGILRRNRPGKTSRRRRTRRSQVAKAYDE"
 2735..2977
 /gene="cut5a"
 /locus_tag="RV3724a"
 2735..2977
 /gene="cut5a"
 /locus_tag="RV3724a"
 /EC_number="3.1.1.-"
 /function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE STRUCTURE OF PLANT CUTICLE)."
 /note="RV3724a, (MTV025.072), len: 80 aa. Probable cut5a, truncated cutinase precursor (EC 3.1.1.-), similar to N-terminal end of others e.g. Q9K87 SERINE ESTERASE CUTINASE from Mycobacterium avium (220 aa), FASTA scores: opt: 202, E(): 1.5e-06, (56.45% identity in 62 aa overlap); Q9XB09|RVND2-RV1758 PROTEIN (FRAGMENT) from Mycobacterium bovis BCG (143 aa), FASTA scores: opt: 200, E(): 1.5e-06, (61.4% identity in 57 aa overlap); and Q00298|CUT1 BOTCI|CUT9A CUTINASE PRECURSOR from Botrytis cinerea (Botryotinia fuckeliana) (202 aa), FASTA scores: opt: 108, E(): 2.2, (40.4% identity in 52 aa overlap). Also highly similar to others from Mycobacterium tuberculosis e.g.
 006318|CUT1_MYCTU|RV3451|MT3557|MTCV13E12.04 PROBABLE CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 189, E(): 1.2e-05, (58.0% identity in 50 aa overlap); 050664|CUT2_MYCTU|RV3301|MT2356|MTCY339.08c PROBABLE CUTINASE PRECURSOR (219 aa), FASTA scores: opt: 172, E(): 0.0015, (59.2% identity in 49 aa overlap); 006793|RV1758|MTCY28.24|Z95890 HYDROTHERICAL 17.9 kDa PROTEIN (174 aa), FASTA scores: opt: 641, E(): 2.7e-29, (57.2% identity in 166 aa overlap); 006319|RV3452|MTV13E12.05; and U00015.11 from Mycobacterium leprae. BELONGS TO THE CUTINASE FAMILY. Rest of cutinase ORF continues as RV3724b|CUT5B, frameshifting could occur near position 4169668. Sequence has been checked but no errors found."
 /codon_start=1
 /transl_table=11
 /product="PROBABLE CUTINASE PRECURSOR [FIRST PART] CUT5a"
 /protein_id="CAE5633.1"
 /db_xref="GI:38490372"
 /translation="MDVIRMAKRLAIVAGTAAATTTGGLSAHVPMVSAEPCDVEVVFARGGEPPGIGSGGLFVDALRFPGRKQVTRGLRR"
 2874..3437
 /gene="cut5b"
 /locus_tag="RV3724b"
 2874..3437
 /gene="cut5b"
 /EC_number="3.1.1.-"
 /function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE STRUCTURE OF PLANT CUTICLE)."
 /note="RV3724b, (MTV025.072), len: 187 aa. Probable cut5b, truncated cutinase (EC 3.1.1.-), similar to C-terminal end

of others e.g. Q9XB09|RV2D-RV1758 PROTEIN (FRAGMENT) from Mycobacterium bovis BCG (143 aa) FASTA scores: opt: 33.5, E(): 3.4e-12 (53.25% identity in 92 aa overlap); Q9XB87 SERINE ESTERASE CUTINASE from Mycobacterium avium (220 aa), FASTA scores: opt: 25.1, E(): 2.5e-07, (44.05% identity in 168 aa overlap). Also similar to proteins from Mycobacterium tuberculosis e.g. O06793|RV1758|MYCY28.24 HYPOTHETICAL 17.9 KDA PROTEIN (174 aa), FASTA scores: opt: 64.1, E(): 2.5e-29, (57.25% identity in 166 aa overlap); O06319|RV3452|MYCY1312.05 HYPOTHETICAL 23.1 KDA PROTEIN (225 aa), FASTA scores: opt: 38.5, E(): 7.5e-15, (46.65% identity in 165 aa overlap); O06318|CUT3|MYCTU|RV3451|MT3557|MYCY1312.04 PROBABLE CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 30.7, E(): 1.9e-10, (40.7% identity in 167 aa overlap); O10837|CUT1|MYCTU|RV1984C|MT2037|MYCY39.35 PROBABLE CUTINASE PRECURSOR (217 aa), FASTA scores: opt: 26.1, E(): 6.7e-08, (50.9% identity in 169 aa overlap); etc; and U00015.11 from Mycobacterium lepra. 5'-end of gene is RV3724A|CUT5A; frameshifting may occur near position 4169668. TBPase score is 0.918."

/codon_start=1
/transl_table=1
/product="PROBABLE CUTINASE [SECOND PART] CUT5B"
/protein_id="CAE55633.1"
/db_xref="GI:38490373"
/translation="MAPGSHLVLAASEDCSTHCVSQVGAKSLGVAVNYPASNDPAS
SDPKVTVDIGRIDAGSHISQSMASCPOTROYLGYSQGAAYAVGVTSVAVPAPVQA
VPRMAPEVNHNVAATVTFGAPSAQFQGYCAPPAIAGLPYQPTLQLCADGDSICGD
GNSPVAHGLVAVNGVQGANPAASRL"

Query Match 100.0%; Score 20; DB 1; Length 244800;
Beat Local Similarity 100.0%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TCGAAGGTGGCAATCGTG 20
|||||
Db 226523 TCGAAGGTGGCAATCGTG 226542

RESULT 6
BX248347
LOCUS 278492 bp DNA linear BCT 06-MAY-2004
DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment 14/14.
ACCESSION BX248347 BX248333
VERSION BX248347.1 GI:31620456
KEYWORDS complete genome.
SOURCE Mycobacterium bovis AF2122/97
ORGANISM Mycobacterium bovis AF2122/97
Corynebacteriales; Actinomycetales;
tuberculosis complex.
1
Garnier,T., Biglmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
Pryor,M., Duhoy,S., Grondin,S., Lacroix,C., Monsempe,C., Simon,S.,
Harriss,B., Acklin,R., Doggett,D., Mayes,R., Keating,L.,
Wheeler,P.R., Parkhill,J., Barrrell,B.G., Cole,S.T., Gordon,S.V. and
Hewinson,G.
The complete genome sequence of Mycobacterium bovis
Unpublished
2 (bases 1 to 278492)
Garnier,T.
Direct Submission
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
Bacterienne Institut Pasteur 28,rue du Dr Roux 75724 PARIS cedex
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK. P4 Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28

FEATURES SOURCE

gene
CDS

gene
CDS

gene
CDS

rue du Docteur Roux, 75724 Paris Cedex 15, France

1..278492
/organism="Mycobacterium bovis AF2122/97"
/mol_type="genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:333413"
complement(240..704)
/locus_tag="MB3713c"
complement(240..704)
/locus_tag="MB3713c"
/note="MB3713c", len: 154 aa. Equivalent to RV3686C.
len: 154 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 154 aa overlap). Hypothetical protein,
similar to other bacterial hypothetical proteins e.g.
Q9X341|SC166.02c from Streptomyces coelicolor (154 aa),
FASTA scores: opt: 42.5, E(): 3.4e-19, (46.1% identity in
154 aa overlap); Q9WZF4|TM0690 from Thermococcus maritima
(149 aa), FASTA scores: opt: 32.6, E(): 3.4e-13, (40.4%
identity in 151 aa overlap); Q9PHU3|CU0573 from
Campylobacter jejuni (147 aa), FASTA scores: opt: 29.0,
E(): 5.1e-11, (36.4% identity in 151 aa overlap); etc.
Also some similarity to upstream
O69654|RV3686C|MTV025.034c CONSERVED HYPOTHETICAL PROTEIN
from Mycobacterium tuberculosis."

/codon_start=1
/transl_table=1
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD95899.1"
/db_xref="GI:31620457"
/db_xref="UniProt/TREMBL:Q7TVX7"
/translation="MAPLKSQRLDLOQAMTODKLTATIRMLAAIQTGEVSGKQA
RELSDDEVIKVLARSRRKGAEIYVONGGEAAATSEARITDELPLPLEGEL
ADVDVTAIAEVAEELGHRPSMKQGLVWKATVIAAGKADGARLSAAVKRL"
704..2059
/locus_tag="MB3714"
704..2059
/locus_tag="MB3714"
/note="MB3714", len: 451 aa. Equivalent to RV3689, len:
451 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.8% identity in 451 aa overlap). Probable conserved
transmembrane protein, with proline rich N-terminus,
similar to Q9RYW6|SC33.17 PUTATIVE INTEGRAL MEMBRANE
PROTEIN from Streptomyces coelicolor (462 aa) FASTA
scores: opt: 73.0, E(): 2.7e-21, (38.1% identity in 412 aa
overlap)."
/codon_start=1
/transl_table=1
/product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN"
/protein_id="CAD95900.1"
/db_xref="GI:31620458"
/db_xref="UniProt/TREMBL:Q7TVX6"
/translation="MHKRYAPQRPKPTETVYIKCTDRDQGGHDERQLRPVMLP
PGYVEPPVAPAGYAPAGYPPYAPATPPYGGPAGPSPGPGYGGPGLYAPAPG
GVVPLQIISIALGPMPTAFGDIYGBDELDTLVAGGWSAFASLVSADGVLAS
AAALTVIGRAVFGPIVGEAMAVRGLALFGALAEAGVAVGLAVIISGVA
GAANBAALALGFPILVVGSLAVLVLLFALVLLLEPLVLEATRFALVRVAG
FMVLTGIRLTVLVGVGNALAPFMIIVGIVAVVAVASDGSYMRVLGATLSAIGVT
IGQIVTAPFSAAGVVLVYTDRIIAEAFDLVQLGLEAGPAGAPVESTNMLWLRP
F"
2086..2739
/locus_tag="MB3715"
2086..2739
/locus_tag="MB3715"
/note="MB3715", len: 217 aa. Equivalent to RV3690, len:
217 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 217 aa overlap). Probable conserved
membrane protein, similar to Q9KYW5|SC33.18 PUTATIVE
INTEGRAL MEMBRANE PROTEIN from Streptomyces coelicolor
(231 aa), FASTA scores: opt: 41.9, E(): 1.5e-19, (36.0%
identity in 211 aa overlap). Equivalent to AAK48159 from
Mycobacterium tuberculosis strain CDC1551 (233 aa) but

```
shorter 16 aa."
/codon_start=1
/translate_table=11
/product="PROBABLE CONSERVED MEMBRANE PROTEIN"
/protein_id="CAD95901.1"
/db_xref="GI:31620459"
/translation="MPSIDIREAHAAOAEKDIYKPSITKELTDWIDEOLYRI
LEKSSITGCGFTITVLLIMIAVTAQVARRRTRRGDDYLPFAGQUTAAQHR
STASRYAEKGMMAIRLRQVARELEETGMLNPAQRATVELSDGEVPLHAGS
LTQATAFNDVYTGRRPQTQAYQMIADLDHLRSRPAVVSAYQHPAPVPSMAQVR"
2865..3866
/locus_tag="MB3716"
/locus_tag="MB3716"
/locus_tag="MB3716"
/locus_tag="MB3716", len: 333 aa. Equivalent to RV3691, len:
333 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 333 aa overlap). Conserved
hypothetical protein, similar to O9KYM4|SCE33.19 PUTATIVE
SECRETED PROTEIN from Streptomyces coelicolor (387 aa)
FASTA scores: opt: 481, E(): 6e-23, (36.6% identity in 358
aa overlap). Equivalent to AKK48160 from Mycobacterium
tuberculosis strain CDC1551 (381 aa) but shorter 48 aa."
/codon_start=1
/translate_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD95902.1"
/db_xref="GI:31620460"
/db_xref="UniProt/TREMBL:Q7TVX4"
/translation="MAPASTSTGSHALATLGNHGVVVVADSIADVEAARPSLL
LVAOTQYLVNADLRLAKAPGDLIVATRTLRALPOLRIAAFPNSQNTLR
EANRAGSYQMGPSDTQATGDLVITSCGALVRRARAGRTIVVSGSNPTNGSLP
AGNALAMNLGNRPRIWYAPDHIEGMSPPSSSDIPENVHTITQMLVITLVA
LMKRRIPFLVAELPVIYRASLETVEGRLRSRRADRADARLALQLRRRLG
VGAQAPAVAVTTIAQRSKADPPFVAYHLFGPAPATDNDLQLARALDIRQVYHS"
3863..4939
/locus_tag="MB3717"
/locus_tag="MB3717"
/locus_tag="MB3717"
/locus_tag="MB3717"
/locus_tag="MB3717", len: 358 aa. Equivalent to RV3692,
len: 358 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 358 aa overlap). Probable moxR2,
methanol dehydrogenase regulatory protein, highly similar
(generally longer at N-terminus) to O9KYM3|SCE33.20
PUTATIVE REGULATORY PROTEIN from Streptomyces coelicolor
(329 aa), FASTA scores: opt: 1523, E(): 4.2e-74, (70.9%
identity in 330 aa overlap); O92538|SC982.21c PUTATIVE
REGULATORY PROTEIN from Streptomyces coelicolor (332 aa)
FASTA scores: opt: 1008, E(): 1.1e-46, (50.8% identity in
313 aa overlap); Q9U267|MOXR-3|PAB0848 METHANOL
DEHYDROGENASE REGULATORY PROTEIN from Pyrococcus abyssi
(314 aa), FASTA scores: opt: 989, E(): 1.1e-45, (50.65%
identity in 302 aa overlap); Q9AANI|CC0566 MOXR PROTEIN
from Caulobacter crescentus (323 aa), FASTA scores: opt:
988, E(): 1.3e-45, (52.3% identity in 306 aa overlap);
etc. Also similar to O53170|MTV007.26|MOXR|RV1479 from
Mycobacterium tuberculosis (377 aa); and
O07392|AF002133.6|MOXR from Mycobacterium avium (309 aa).
Also high similarity with several hypothetical bacterial
proteins."
/codon_start=1
/translate_table=11
/product="PUTATIVE METHANOL DEHYDROGENASE TRANSCRIPTIONAL
REGULATORY PROTEIN MOXR2"
/protein_id="CAD95903.1"
/db_xref="GI:31620461"
/db_xref="UniProt/TREMBL:Q7TVX3"
/translation="HTGSASNPQAPPTTPGAEIPYPOQAGAPTAAPSGHPHRAE
AESADALALRAEYAKAVAGQDVTSLVITALCRGVLLLEGVGAKTILIVAMSA
ALILEFRKRVQTPDLMGDTVSLVYDRTAEFVRPGVFTNLADIEIKRTPKTO
AALIEAMERQVSVEGEPKPLNPPIVAATONPIEYEGTYQLPEAQDLRFLKINVTL
```

```
PARDEIALLDRHAHPDPRDLASINPAGAEIAGREAVRHUVANVUCYIVDI
GATRSSPALQCVSPRGATLALGTRASWAMLSGRDYVTPDDVKAMARPTLRVWLRP
EAELEGATPDDVGLGILASVPVR"
5073..6395
/locus_tag="MB3718"
/locus_tag="MB3718"
/locus_tag="MB3718"
/locus_tag="MB3718", len: 440 aa. Equivalent to RV3693, len:
440 aa (alternative start at 41910), from Mycobacterium
tuberculosis strain H37Rv, (100.0% identity in 440 aa
overlap). Possible conserved membrane protein, similar to
O9KYM2|SCE33.21 PUTATIVE LIPOPROTEIN from Streptomyces
coelicolor (436 aa), FASTA scores: opt: 875, E(): 3.3e-46,
(56.25% identity in 448 aa overlap); Q9AANI|CC0567
HYPOTHETICAL PROTEIN from Caulobacter crescentus (437 aa),
FASTA scores: opt: 355, E(): 2.3e-14, (30.9% identity in
450 aa overlap); P73233|SLR2013 HYPOTHETICAL 48.5 KDA
PROTEIN from Synecocystis sp. strain PCC 6803 (435 aa),
FASTA scores: opt: 340, E(): 1.9e-13, (29.7% identity in
438 aa overlap); etc. Equivalent to AKK48162 from
Mycobacterium tuberculosis strain CDC1551 (475 aa) but
shorter 35 aa. Also similar to other hypothetical proteins
from Mycobacterium tuberculosis; MTV014_7; MTV007_27; and
Query Match 100.0% Score 20; DB 1; Length 278492;
Beet Local Similarity 100.0%; Pred. NO. 54;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCGAAGTGTGGCAATCGTG 20
DB 260216 TCGAAGTGTGGCAATCGTG 260235
RESULT 7
BX248359 349895 bp DNA linear BCT 06-NOV-2003
LOCUS BX248359
DEFINITION Corynebacterium diptheriae gravis NCTC13129, complete genome;
segment 6/8
ACCESSION BX248359 BX248353
VERSION BX248359.1 GI:38200531
KEYWORDS complete genome.
SOURCE Corynebacterium diptheriae
ORGANISM Corynebacterium diptheriae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE
1 (bases 1 to 349895)
Cerdeno-Tarraga A.M., Efratiou A., Dover L.G., Holden M.T.G.,
Pallen M., Bentley S.D., Besra G.S., Churche C., James K.D., De
Zorja A., Chillingworth T., Cronin A., Dowd L., Fellwell T.,
Hamlin N., Holroyd S., Jagels K., Kettle S., Quail M.A.,
Rabinowitsch E., Rutherford K., Thomson N.R., Unwin L.,
Whitehead S. and Barrall B.G. Parkhill J.
The complete genome sequence and analysis of Corynebacterium
diptheriae NCTC13129
Nucleic Acids Res. 31 (22), 6516-6523 (2003)
JOURNAL PUBMED
14602910
REFERENCE
2 (bases 1 to 349895)
Cerdeno-Tarraga A.M.
Direct Submission
Submitted (03-OCT-2003) Cerdeno-Tarraga A.M., submitted on behalf
of the Pathogen Sequencing Unit, Sanger Institute, Wellcome Trust
Genome Campus, Hinxton, Cambridge CB10 1SA E-mail:
amct@sanger.ac.uk
Location/Qualifiers
1..349895
/organism="Corynebacterium diptheriae"
/mol_type="genomic DNA"
/strain="NCTC13129"
/db_xref="taxon:1717"
/notice="biotype gravis"
91..648
/locus_tag="DIP1698"
91..648
CDS
gene
CDS
```

```
/locus_tag="DIP1698"
/note="Similar to Corynebacterium glutamicum thiamine
biosynthesis protein X Thx SW:THX_CORGU (P42461) (33
aa) fasta scores: E(): 2.5e-12, 38.09% id in 126 aa"
/codon_start=1
/transl_table=1
/product="putative thiamine biosynthesis related protein"
/protein_id="CAE50227.1"
/db_xref="GI:38200532"
/translation="MTIRIRSLTTAAVIALGAIACSPPHQVSDTVDRATSK
VAQPSASASAMAPSDDKNAETPASATENLTNGONTVELSGNPEVGYAAICNAT
TPICGAPRLCTGGADTSAQAMIKADGTGTALTNDGTATVNLTAATGGQVNCIKDK
CVIKFGDHTGEGRDTEVPTFAS"
91..189
/locus_tag="DIP1698"
/note="Signal peptide predicted for DIP1698 by SignalP 2.0
HMM (Signal peptide probability 1.000) with cleavage site
probability 0.728 between residues 33 and 34;
signal-peptide site"
337..360
/locus_tag="DIP1698"
/note="ScanRegExp hit to P500687, Aldehyde dehydrogenases
glutamic acid active site."
complement(653..700)
/note="Score 63: 21/21 (100%) matches, 0 gaps"
/gene="dnaG"
/locus_tag="DIP1699"
complement(697..2595)
/gene="dnaG"
/locus_tag="DIP1699"
complement(697..2595)
/BC_number="2.7.7.-"
/note="Similar to Mycobacterium tuberculosis DNA primase
DnaG or RV2343c or MT2408 or MTCY98.12c SW:PRIM_MYCTU
(P95239) (639 aa) fasta scores: E(): 1.4e-146, 59.11% id
in 631 aa, and to Escherichia coli DNA primase DnaG or
DnaP or ParB or B3066 or Z4419 or EC63949 SW:PRIM_ECOLI
(P02923) (581 aa) fasta scores: E(): 2e-38, 32.27% id in
505 aa"
/codon_start=1
/transl_table=1
/product="DNA primase"
/protein_id="CAE50228.1"
/db_xref="GI:38200533"
/translation="MAGKGRIPDSGLAIRRTPIREIVGVOLKPAQDSLKGLSP
FKREKTPSFHVRPNHGFHRCSTGKGVPSFLMEMEGLTPEAVEVCAEKIQINY
QGTTGARREEPGRKRLIEANRVAHQPYRQQLTPEAAVARQFLDGFQOHYGF
BCGYAPBGMDTMTKFMLEKGFKELEAAGSTWGRQPIDRFHRLIMPITKSYGDV
IGFARKLFDDDKLGYMNTPEETLKYKSKVLFGIDLAKKSIAHQHVVVEGYTDM
AMHAAGTTTAAACGTAEGEHLIOLIRYMLDDAVFASGLYTFEGDGAQKAMRAP
EGDOKFGOSYVAAPDGMPCDLRLAGDAAVDLVQRPVMEFIIRTLISEPLD
SVEGRLQLRAVPVASTIRDPALKTETARLLAGVCGPDPEVILQYRAERRPKTO
PRQRATRIDNSQPMATQSMVMDREKPEPLMPQREALKIALQTPHAGVFPDLPD
DAFTESYAAVRKAIVAAGCSDAIRGSTWLDADVADNYDVAGKSLVSELAVEHILD
EMGLAHTDTTLARMEQRIGNQIALKGAMQMRPMSBDMOAYNSLFDVALAEQTRR
ELKXIFK"
complement(1546..1806)
/locus_tag="dnaG"
/locus_tag="DIP1699"
/note="HMPfam hit to PF01751, Toprim domain"
complement(1570..1806)
/gene="dnaG"
/locus_tag="DIP1699"
/note="HMMsmat hit to SM00493, No definition"
complement(2281..2577)
/gene="dnaG"
/locus_tag="DIP1699"
/note="HMPfam hit to PF01807, CHC2 zinc finger"
complement(2311..2478)
/gene="dnaG"
/locus_tag="DIP1699"
/note="BlastP xdb hit to PD002988, PD002988"
complement(2317..2481)
```

```
/gene="dnaG"
/locus_tag="DIP1699"
/note="HMMsmat hit to SM00400, zinc finger"
complement(2606..4483)
/gene="glnS"
/locus_tag="DIP1700"
complement(2606..4483)
/gene="glnS"
/locus_tag="DIP1700"
complement(2606..4483)
/BC_number="2.6.1.16"
/note="Similar to Streptomyces coelicolor
glucosamine--fructose-6-phosphate aminotransferase
[isomerizing] GlnS or SC664.18 SW:GlnS_STRO (O86781) (614
aa) fasta scores: E(): 2.7e-154, 65.86% id in 624 aa, and
to Escherichia coli glucosamine--fructose-6-phosphate
aminotransferase [isomerizing] GlnS or B3729 SW:GlnS_ECOLI
(P71769) (608 aa) fasta scores: E(): 1e-90, 45.71% id in
630 aa"
/codon_start=1
/transl_table=1
/product="glucosamine--fructose-6-phosphate
aminotransferase [isomerizing]"
/protein_id="CAE50229.1"
/db_xref="GI:38200534"
/translation="MCGIVGFVQRTSVDPDYPALDVLEGLRLEYGDSAGVAV
ADGAVSFRKAGKQVALQLELETSPMPQSCIGIHTMTWATGPTDANAPHVVDGK
LAVVNGIINFPALGKSELGFGHNFSVSEDTVAATLLGHIENNANKDLTRAMQLT
CORLEGAFTLAIHAETPRDIVAARNSPVIQEGENFLGSDVSGFIDTKNAVEN
DNDQIVTTITDGHITDPQGNHAGKFPVVEPMQAQAEKGFVEFMKEITHQPAAYR
DTLMGRPDSGKLTDLRLIDESTLRSIDKIYIACGTAAYAGVAYALIEHMKRIPT
EVELAHSEFRYRDPVINEKTVLTLSQSGEMTDLMAARHAEQAKVIALCNTGSSIT
PRESDACTLYHAGPEIAVASTKAFIAQTITATYLLGLAQLRGMFPADENAVAGELR
TIPDKVAVLVDGVDQVTLADMKDAITSVLFLGRHYGFPVLALEGALIKELAYLAHE
TPAGALGKHGPIILIEEGOPVFPVPRGRSDSLAHKAVNSIQEVRARGAITVIABE
GDPAVEAYAHNIIIRIPAPILMQPLATVPLQIFACGVAAKAGFDVQPNLASVTV
E"
complement(2651..3073)
/gene="glnS"
/locus_tag="DIP1700"
/note="HMPfam hit to PF01380, SIS domain"
complement(3188..3592)
/gene="glnS"
/locus_tag="DIP1700"
/note="HMPfam hit to PF01380, SIS domain"
complement(3905..4480)
/gene="glnS"
/locus_tag="DIP1700"
/note="HMPfam hit to PF00310, Glutamine amidotransferases
class-II"
complement(4466..4483)
/gene="glnS"
/locus_tag="DIP1700"
/note="ScanRegExp hit to P500443, Glutamine
amidotransferases class-II active site."
4574..5032
/locus_tag="DIP1701"
/locus_tag="DIP1701"
complement(4574..5032)
/locus_tag="DIP1701"
/note="Similar to Streptomyces aureofaciens
guanylyl-specific ribonuclease S43 precursor RNase3
SW:RNS3_STRAU (P30289) (141 aa) fasta scores: E():
1.7e-10, 42.14% id in 140 aa"
/codon_start=1
/transl_table=1
/product="putative ribonuclease"
/protein_id="CAE50230.1"
/db_xref="GI:38200535"
/translation="MSGAKSTPYHKKPSLVAAIAGLAIAVAGYFGIQPSQDSFQQA
PATSAIRHEDCEVSESLPPQVKDITADITAGGFVDNDGVFNGFEGHLPQRDRN
FRRYVETRPGLRRHGERRIITGGSKTSPQMWYTTDDHYSCFIRPAH"
4574..4705
/locus_tag="DIP1701"
/note="Signal peptide predicted for DIP1701 by SignalP 2.0
```


HMM (Signal peptide probability 1.000) with cleavage site probability 0.513 between residues 44 and 45:
 misc_feature
 signal-peptide site"
 4616..4675
 /locus_tag="DIP1701"
 /note="1 probable transmembrane helix predicted for DIP1701 by TMHMM2.0"
 4754..5026
 /locus_tag="DIP1701"
 /note="TMHMM2am hit to PF00545, ribonuclease."
 5079..5261
 /locus_tag="DIP1702"
 5079..5261
 /locus_tag="DIP1702"
 /note="No significant database matches"
 /codon_start=1

Query Match 90.0%; Score 18; DB 1; Length 34895;
 Best Local Similarity 100.0%; Pred. No. 5e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GAAGTGGGCAATCGTG 20
 |||||
 Db 348977 GAAGTGGGCAATCGTG 348994

RESULT 8
 AC135920/c
 LOCUS
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 5 clone
 OSUNBA0015G13, complete sequence.
 AC135920
 AC135920.2 GI:44151402
 HTG.
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrharioideae; Oryzoae; Oryza.
 1 (bases 1 to 143966)
 Chow, T.-Y., Hsiang, Y.-I. C., Chen, C.-S., Chen, H.-H., Liu, S.-M.,
 Chao, Y.-T., Chang, S.-J., Chen, H.-C., Chen, S.-K., Chen, T.-R.,
 Chao, Y.-L., Cheng, C.-H., Chung, C.-I., Han, S.-Y., Hsiao, S.-H.,
 Hsiung, J.-N., Hsu, C.-H., Huang, J.-J., Kau, P.-I., Lee, M.-C.,
 Lue, H.-L., Li, Y.-F., Lin, S.-J., Lin, Y.-C., Wu, S.-W., Yu, C.-Y.,
 Yu, S.-W., Wu, H.-P. and Shaw, J.-F.
 Oryza sativa BAC OSUNBA0015G13 genomic sequence
 Unpublished
 2 (bases 1 to 143966)
 Chow, T.-Y. and Hsiang, Y.-I. C.
 Direct Submission
 Submitted (25-OCT-2002) Institute of Botany, Academia Sinica, 128,
 Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
 3 (bases 1 to 143966)
 Chow, T.-Y.
 Direct Submission
 Submitted (27-FEB-2004) Institute of Botany, Academia Sinica, 128,
 Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
 4 (bases 1 to 143966)
 Chow, T.-Y. and Hsiang, Y.-I. C.
 Direct Submission
 Submitted (01-AUG-2004) Institute of Botany, Academia Sinica, 128,
 Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
 On Feb 27, 2004 this sequence version replaced gi:24371315.
 Genes were predicated from the integrated results of the following:
 BLASTN.0, BLASTX.0, GENSCAN (Chris Burge,
 http://genes.mit.edu/GENSCAN.html), Egenesh
 (http://www.softberry.com/), Glimmer
 (http://www.tigr.org/softlab/glimmer/glimmer.html), TWINSKAN
 (http://genes.cs.wustl.edu/) and Genesplitter
 (http://www.tigr.org/tdb/Genesplitter/index.shtml). The sequence was
 searched against the Swiss-Prot+TrEMBL protein database, the NCBI
 Plant EST database, the TIGR Rice Gene Index and the rice

full-length cDNA database (KOME,
 http://cdna01.dna.affrc.go.jp/cDNA/). Annotated genes are named to
 indicate the level of evidence for their annotation. Genes with
 similarity to other proteins are named after the database hits.
 Genes without significant peptide similarity but with EST
 similarity are named as unknown proteins. Genes with protein or
 EST similarity, that are predicted by more than two gene prediction
 programs over most of their length are annotated as hypothetical
 proteins. This clone overlaps with R0708D12 (accession # AC130732)
 and OSUNB00115F21 (accession # AC135918).
 Location/Qualifiers

FEATURES
 source
 1..143966
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="Genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="5"
 /clone="OSUNBA0015G13"
 3206..5357
 /gene="OSUNBA0015G13.1"
 /note="hypothetical protein"
 join(3206..3386,3466..3731,3918..3959,4266..4449,
 4684..4883,4977..5357)
 /gene="OSUNBA0015G13.1"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAT85107.1"
 /db_xref="GI:50878332"
 /translation="MSSRYGVSESDVLAHVMNEMLTRDEAIHWLLEDFELMESQRFD
 DKLRLLQMFVEKERMPTKSTPGSAATTKAVLNGVLPKPERVFPCTSLFAPSII
 SAARATLPSIETEAEGDMQVEEETDLHDCAKVELKORADSLFVBLKARPP
 PTFRTSTPSIVAAQKLLLETNORTLITRPSIEBKVAVGLKDKIRLEDVDFMKIL
 GLHKEVVPKSLIKNMGRAVLSIKAWYDERTMTLESTKKIKENKIVYHNIT
 ILVDPGKNDTLGERVANYMQLQTHMDPEGTGNHRLKESKNGKRGGLGTQJG
 RMALVMFQSWPKQALAAIYEQQQAALGIRKLESKAAASATTAATVAGLGTLL
 DLILYPSFSFYCELTITTFYSVCIL"
 13801..15647
 /gene="OSUNBA0015G13.2"
 /note="putative metallothionein"
 join(13801..13850,13996..14043,15548..15647)
 /gene="OSUNBA0015G13.2"
 /note="similar to rice EST CF337510"
 /codon_start=1
 /product="putative metallothionein"
 /protein_id="AAT85108.1"
 /db_xref="GI:50878333"
 /db_xref="UniProt/TrEMBL:O22376"
 /translation="MSDKCGKCDACDSQCVKGTSGYGVVIVDAEKSHFENAEVGYE
 ENDKCKCTTSCSAGCNCGK"
 complement(25750..26234)
 /gene="OSUNBA0015G13.3"
 /note="unknown protein"
 complement(join(25750..26112,26172..26234))
 /gene="OSUNBA0015G13.3"
 /note="similar to rice EST AK107744"
 /codon_start=1
 /product="unknown protein"
 /protein_id="AAT85109.1"
 /db_xref="GI:50878334"
 /db_xref="UniProt/TrEMBL:O9AW66"
 /translation="MARLIGIVAAALAAVAVATLVPAAHHPPPRHLLLVVRQGA
 VLSGGGRRRAAPAGAGESEADPEVROLPGREGRGVQOARPLLPELRVAAGESL
 PRLRLRYPLLPQELTMMVVTSLSIQIPFIRPIH"
 complement(36352..36954)
 /gene="OSUNBA0015G13.4"
 /note="hypothetical protein"
 complement(36352..36954)
 /gene="OSUNBA0015G13.4"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAT85110.1"
 /db_xref="GI:50878335"
 /db_xref="UniProt/TrEMBL:Q9SYMS"

gene
 CDS

gene
 CDS

gene
 CDS

gene /translation="MAPPAASVSAADAAASYTTTPNITLITGAGFIASHVANRLI
RPHRSYRVATLADKDDYCATLNNAPAMASGNPAPFVPGDVAADLVHILAEINIV
MHPAAHTVVDNSFEFTKNVLTGHVLEPCCRKAAPPAPSAASSTAPTRSTATLPRR
ATPRRGSGSPPTNLKSGQRHLAVSFISITRYNGLKL"
/gene="OSUNBA0015G13.5"
/note="unknown protein"
/join(41341..41706,41783..42475)
/gene="OSUNBA0015G13.5"
/note="similar to rice EST CF323965, AK064593"
/codon_start=1
/product="unknown protein"
/protein_id="AAT8511.1"
/db_xref="GI:50878336"
/db_xref="UniProt/TREMBL:O81M04"
/translation="MDIGESKGGKGGKSSNLAAGRAPINWPLALSEFLIDWYIEK
KLQLEPKAVIKIHHTACTLAINSRYGTYTVDOVQHHYRHKEMWSLVAAHLNBSGN
GMDETTKMLTLSQATLDALLINDRGLSKPQFPDKQLFSGSGSADGAFMEDSSAA
DPDEDDDDKDFDLMDSTYADTVPGQEDFDKLESDDCKEVAALSAATIQVSS
TNVLDKPKKKNKCGKPKTLPOSHNDCKNTKATQVHSDVDVLIINNTLVGIK
MTLEKPIOTVAPODPNIPLMMDIKKIALDPDCLKRVGHLCKPELQANRSLISMGE
YLEHWITKFLSGDDPNLIG"
complement(53558..54387)
/gene="OSUNBA0015G13.6"
/note="hypothetical protein"
complement(join(53558..53812,54340..54387))
/gene="OSUNBA0015G13.6"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAT8511.1"
/db_xref="GI:50878337"
/translation="WHYKQBAEPSTGLAMPAPASPGSGSGESSIEDTADROY
TPCRRNGLKKAYELSMCDAREVALITYSSRGRUYEISNNRYCIPLPFSSRNIN"
complement(60550..64519)
/gene="OSUNBA0015G13.7"
/note="hypothetical protein"
complement(join(60550..60861,61654..62031,63576..63754,
63866..64134,64329..64519))
/gene="OSUNBA0015G13.7"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAT8511.1"
/db_xref="GI:50878338"
/db_xref="UniProt/TREMBL:Q7XRJ3"
/translation="MCVRYALNVTINKKYSLEPIYDLFPOILKGAITYFLDRPITV
SPATNSRRYPODSLHPIRHHQIAEFSKCFMLSKYFLDHYTFPGVAVDPAIVE
AVTKWRPITDHSILGLAFRRFTEFLKIAMPWTQLKKEEFMTLCCLELRVP
SGFGLRPNAGACGGLRLATATATAMQLPHQPRIGRIACSENNALASHREPIFLR
KHTRLARKHSGQSDASTWIGPTSQVPHGKHCHQCPVSPAKSEARWGP
WFGTRGSAEPRLAFLAFHYAVDDIPMTVSVIPETWMTGQERDALMWLKH
LHPRGHKVSSSHPRLSNDLGGESATSDPRGKDRPRHGLIAEHKRESVWL
HGRRTWHLNMTGICPKDPKIKLILBAHKSQYSHPGQHDPFRKG"
complement(66976..77822)
/gene="OSUNBA0015G13.8"
/note="putative MADS box transcription factor"
complement(join(66976..66979,67483..67544,67729..67770,
67872..67913,68013..68112,68207..68268,68808..68889,
76958..77217,77775..77822))
/gene="OSUNBA0015G13.8"
/note="similar to rice EST CR393425, AK111723"
/codon_start=1
/product="putative MADS box transcription factor"
/protein_id="AAT8511.1"
/db_xref="GI:50878339"
/db_xref="UniProt/TREMBL:Q8GTP4"
/translation="WHYKQBAEPSTGLMPEPAPVASPGSGSGSGSVAEKIGS
RGKIEIKRIENTNDFECCRSGGLTKKAYELSVLCDAEVALVFSRGRALYEYSN
SVKETIERKXANDSTNSATVAETINAOHYOEAAKIQOITNLINNSRLLVGDNITT
MMHRELKQLEGLDGLGKIRAKRNEILCAELFVWQRRETELQNDNMLKSVASER
GLQTYMMGSAISSE"
complement(88454..96189)
/gene="OSUNBA0015G13.9"
/note="unknown protein"

CDS
complement(join(88454..90352,95721..96066,96167..96189))
/gene="OSUNBA0015G13.9"
/note="similar to rice EST AK069504"
/codon_start=1
/product="unknown protein"
/protein_id="AAT8511.1"
/db_xref="GI:50878340"
/db_xref="UniProt/TREMBL:Q9AW18"
/translation="MAPSFSTRVAAAMRGTAASRHPALSAAGDFPPPIRRFRMWVD

Query Match 87.0%; Score 17.4; DB 8; Length 143966;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 1 TCGAAGTGGGCAATGCT 19
Db 91684 TCGAAGTGGGCAATCAT 91666

RESULT 9
AC122848
LOCUS
DEFINITION
MUS musculus BAC clone RP23-166K13 from chromosome 18, complete
sequence.
AC122848
AC122848.3 GI:38564454
HTG.
MUS musculus (house mouse)
SOURCE
ORGANISM
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Tomlinson, C., Cotton, M., Bielicki, L. and Haskenson, W.
The sequence of Mus musculus BAC clone RP23-166K13
Unpublished (2001)
2 (bases 1 to 204254)
Wilson, R.
Sequencing of Mus musculus
Unpublished (2001)
3 (bases 1 to 204254)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (03-SEP-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 204254)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (28-NOV-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 204254)
Wilson, R.K.
Direct Submission
Submitted (28-NOV-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
6 (bases 1 to 204254)
Wilson, R.
Direct Submission
Submitted (01-JAN-2004) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 28, 2003 this sequence version replaced gi:12657947.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: MUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.wustl.edu
----- Summary Statistics
Center project name: M_BA0166K13

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-23 BAC library has been constructed by Kazutoyo Osegawa and Minako Tateno in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC132427 and AC125110.

FEATURES**source**

1.204254

Location/Qualifiers

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="18"

/map="18"

/clone="RP23-166K13"

/clone_lib="RPCI-23"

247..376

/rpt_family="Alu"

756..808

/rpt_family="MIR"

1826..2049

/rpt_family="MER1_type"

2425..2601

/rpt_family="B2"

2758..2898

/rpt_family="Alu"

3672..3897

/rpt_family="ERVK"

4709..4870

/rpt_family="L1"

4972..5076

/rpt_family="Alu"

5144..5317

/rpt_family="B2"

5703..6147

/rpt_family="MALR"

6489..6887

/rpt_family="MALR"

7826..8220

/rpt_family="MALR"

8817..8898

/rpt_family="MALR"

9632..9778

/rpt_family="Alu"

11045..11290

/rpt_family="L1"

12803..12897

/rpt_family="B4"

13109..13326

/rpt_family="L1"

13358..13470

/rpt_family="L1"

13752..13834

/rpt_family="Alu"

repeat_region 14694..14846
/rpt_family="L2"
repeat_region 14867..14889
/rpt_family="B4"
repeat_region 14890..15026
/rpt_family="Alu"
repeat_region 16119..16197
/rpt_family="B4"
repeat_region 16280..16412
/rpt_family="Alu"
repeat_region 17016..17120
/rpt_family="Alu"
repeat_region 17481..17686
/rpt_family="MER1_type"
repeat_region 19443..19381
/rpt_family="MALR"
repeat_region 20432..20549
/rpt_family="B4"
repeat_region 21468..21577
/rpt_family="MIR"
repeat_region 21738..21817
/rpt_family="L2"
repeat_region 21808..21970
/rpt_family="MIR"
repeat_region 21975..22069
/rpt_family="MALR"
repeat_region 21991..22098
/rpt_family="Alu"
repeat_region 22194..22255
/rpt_family="ERV1"
repeat_region 22373..22471
/rpt_family="ID"
repeat_region 23394..23559
/rpt_family="MER1_type"
repeat_region 23905..24003
/rpt_family="MIR"
repeat_region 24667..24777
/rpt_family="L1"
repeat_region 24778..24827
/rpt_family="B4"
repeat_region 24857..25075
/rpt_family="L1"
repeat_region 25108..25200
/rpt_family="MER1_type"
repeat_region 25824..26052
/rpt_family="B2"
repeat_region 26339..26390
/rpt_family="L2"
repeat_region 26595..26667
/rpt_family="MIR"
repeat_region 26852..26896
/rpt_family="MER1_type"
repeat_region 27760..27818
/rpt_family="L2"
repeat_region 28365..28438
/rpt_family="L1"
repeat_region 30127..30197
/rpt_family="ERV1"
repeat_region 31523..31724
/rpt_family="MIR"
repeat_region 31885..32031
/rpt_family="Alu"
repeat_region 33585..33778
/rpt_family="MIR"
repeat_region 33809..34232
/rpt_family="MALR"
repeat_region 34452..34491
/rpt_family="L1"
repeat_region 34493..34655
/rpt_family="L1"
repeat_region 34822..35154
/rpt_family="L1"
repeat_region 35740..35841

Consensus quality: 174974 bases at least Q30
Consensus quality: 175572 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1242: contig of 1242 bp in length
1243 1342: gap of unknown length
1343 2915: contig of 1573 bp in length
2916 3015: gap of unknown length
3016 4527: contig of 1512 bp in length
4528 4627: gap of unknown length
4628 7912: contig of 3285 bp in length
7913 8012: gap of unknown length
8013 20454: contig of 12442 bp in length
20455 20554: gap of unknown length
20555 27431: contig of 6877 bp in length
27432 27531: gap of unknown length
27532 74998: contig of 47467 bp in length
74999 75098: gap of unknown length
75099 106492: contig of 31394 bp in length
106493 106592: gap of unknown length
106593 147728: contig of 41136 bp in length
147729 147829: gap of unknown length
147829 178455: contig of 30627 bp in length.

FEATURES

source

1..178455
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="7"
/clone="RP43-129D8"
1..1242
/note="assembly_name:Contig42"
1343..2915
/note="assembly_name:Contig45"
3016..4527
/note="assembly_name:Contig48
clone_end:SP6
vector_side:right"
4628..7912
/note="assembly_name:Contig49"
8013..20454
/note="assembly_name:Contig50"
20555..27431
/note="assembly_name:Contig51"
27532..74998
/note="assembly_name:Contig52"
75099..106492
/note="assembly_name:Contig53"
106593..147728
/note="assembly_name:Contig54"
147829..178455
/note="assembly_name:Contig55"

ORIGIN

Query Match 85.0%; Score 17; DB 2; Length 178455;
Best Local Similarity 100.0%; Pred. No. 1.6e+0; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0;

Qy 2 CGAAGTGGCAATCG 18
|||||

Db 85883 CGAAGTGGCAATCG 85889
|||||

RESULT 12
CEY102A5D
LOCUS CEY102A5D 25454 bp DNA linear INV 12-OCT-2004

DEFINITION

Caenorhabditis elegans YAC Y102A5D, complete sequence.

ACCESSION

AL031628
AL031628.2 GI:6425213

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

```

CDS                                complement(join.20322..20673,20849..21515,21771..22002)
                                   /gene="Y102A5D.1"
                                   /standard_name="Y102A5D.1"
                                   /codon_start=1
                                   /product="Hypothetical protein Y102A5D.1"
                                   /protein_id="CAA20973.1"
                                   /db_xref="GI:3880666"
                                   /db_xref="UniProt/TREMBL/Q9XK64"
                                   /translation="MAKDIDAVQDTVNTISAGTGTATTAAGOVTFPAKVPRDVRRVDC
OILQEDIVITLTKEPMNKDEKRECKSTGDPPERGSPLAEYIOIPVKDDGLGVACPPT
TCMTITFLVSLTKLKLMSSDDYTINKUYTSAYEVAEVKQKTGNVDLEDIAELANRYI
RSNGYNCFSRLRDIVIRNLSTVGAPLHKYKVRNKGKLASVQTISMARLYTDADY
HKIQOYLAKELVNKKALKRTFEGEKPKSEVVVRVQQNGSKRPVNRGGYLPLACHRFVH
PALNIINLCWNYNSFDOTRKRGKRGRMKSVSTGGCHLPRLTMNSDETCLKREGKQRS
MLOPHANNENAOVCPPVDRIHLASKSVIPCLCATVMESKNAPMNHIIGDKXTYTPTEL
GHKRXQNOMNKAEPHAPRISTGN"

ORIGIN
Query Match      84.0%; Score 16.8; DB 3; Length 25454;
Best local Similarity 90.0%; Pred. No. 2,1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy       1 TCGAAGCTGGCAATGCTG 20
          ||| |||| |||| |||| |
Db        12033 TCGCAGTGGGCAATCCTG 12052

RESULT_13
AL669827           108424 bp   DNA         linear    ROD 27-APR-2002
LOCUS             Mouse DNA sequence from clone RP23-350D17 on chromosome 11,
DEFINITION        complete sequence.
ACCESSION          AL669827
VERSION            AL669827.5 GI:20339131
KEYWORDS            HTG.
SOURCE              Mus musculus (house mouse)
ORGANISM            Mus musculus (house mouse)
REFERENCE           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS            Mamalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE              1 (bases 1 to 108424)
JOURNAL            Oliver,K.
COMMENT            Direct Submission
                   Submitted (27-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
                   Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerquest@sanger.ac.uk
                   humquery@sanger.ac.uk Clone requests: clonerquest@sanger.ac.uk
                   On Apr 29, 2002 this sequence version replaced gi:20218615.
                   During sequence assembly data is compared from overlapping clones.
                   Where differences are found these are annotated as variations
                   together with a note of the overlapping clone name. Note that the
                   variation annotation may not be found in the sequence submission
                   corresponding to the overlap described above.
                   This sequence was finished as follows unless otherwise noted: all
                   regions were either double-stranded or sequenced with an alternate
                   chemistry or covered by high quality data [i.e., phred quality >=
                   30]; an attempt was made to resolve all sequencing problems, such
                   as compressions and repeats; all regions were covered by at least
                   one plasmid subclone or more than one M13 subclone; and the
                   assembly was confirmed by restriction digest. The following
                   abbreviations are used to associate primary accession numbers given
                   in the feature table with their source databases: Em:, EMBL; Sw:,
                   SWISSPROT; Tr:, TREMBL; Wp:, WormPeP; Information on the WormPeP
                   database can be found at
                   http://www.sanger.ac.uk/Projects/C_elegans/wormpep/RP23-350D17.is
                   from the RPC1-23 Mouse PAC Library
                   constructed by the group of Pieter de Jong.
                   For further details see http://www.chori.org/bacpac/home.htm
                   VECTOR: pBACE6.6.
FEATURES             Location/Qualifiers
                     1..108424
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10090"

```

```

ORIGIN
Query Match      84.0%; Score 16.8; DB 10; Length 108424;
Beet Local Similarity 90.0%; Pred. No. 2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY          1 TCGAAGTGGGCAATCCTG 20
             |||||
DB           62993   TCCGACGTGGGCAGAAGCTG 63012


RESULT 14
LOCUS       ACI36469               147180 bp    DNA         linear     PRI 05-DEC-2002
DEFINITION Homo sapiens chromosome 19 clone RP11-886P16, complete sequence.
ACCESSION   ACI36469
VERSION     ACI36469.2 GI:26050948
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 147180)
            DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submision
JOURNAL     Unpublished
AUTHORS     2 (bases 1 to 147180)
            DOE Joint Genome Institute and Stanford Human Genome Center.
TILE        Direct Submission
JOURNALS    Submitted (02-NOV-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE   3 (bases 1 to 147180)
            DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNALS    Submitted (05-DSC-2002) DOE Joint Genome Institute, 2800 Mitchelll
Drive, Walnut Creek, CA 94598, USA
ON Dec 5, 2002 this sequence version replaced gi:24476050.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Flushing completed at Stanford Human Genome Center
www.sbgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.4.

FEATURES             source
                 .1..147180
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="19"
                /clone="RP11-886P16"

ORIGIN
Query Match      84.0%; Score 16.8; DB 9; Length 147180;
Beet Local Similarity 90.0%; Pred. No. 2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY          1 TCGAAGTGGGCAATCCTG 20
             |||||
DB           137860 TCCAAGTTGGGACAATTCCTG 137879


RESULT 15
LOCUS       ACI36470               156942 bp    DNA         linear     HTG 02-NOV-2002
DEFINITION Homo sapiens chromosome 19 clone RPCI-13_947L3.*** SEQUENCING IN
PROGRESS ***, 3 ordered pieces.
ACCESSION   ACI36470
VERSION     ACI36470.1 GI:24476051
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEPIN.
Homo sapiens (human)

```

```

ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE     1 (bases 1 to 156942)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE         Direct Submission
JOURNAL      Unpublished
REFERENCE     2 (bases 1 to 156942)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE         Direct Submission
JOURNAL      Submitted (02-NOV-2002) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
COMMENT       * NOTE: This is a 'working draft' sequence. It currently
              * consists of 3 contigs. Gaps between the contigs
              * are represented as runs of N. The order of the pieces
              * of the gaps between them are based on estimates that have
              * been provided by the submittor.
              * This sequence will be replaced
              * by the finished sequence as soon as it is available and
              * the accession number will be preserved.
              *
              * 1 10649: contig of 10649 bp in length
              * 10650 10749: gap of unknown length
              * 10750 138483: contig of 127734 bp in length
              * 138484 138583: gap of unknown length
              * 138584 156942: contig of 18359 bp in length.
              -----Genome Center
              Center: Joint Genome Institute
              Center Code: JGI
              Web site: http://www.jgi.doe.gov
              -----
              Project Information
              Center Project Name: 3368425
              Center clone name: RPCI-13_947L3
              -----
              Summary Statistics
              Consensus quality: 155815 bases at least Q40
              Consensus quality: 155971 bases at least Q30
              Consensus quality: 156062 bases at least Q20
              Estimated insert size: 155000; agarose-fp estimation
              Estimated insert size: 156236; sum-of-contigs

estimation    Quality coverage: 1 in Q20 bases; agarose-fp
estimation    Quality coverage: 0.99 in Q20 bases; sum-of-contigs
estimation..  Location/Qualifiers
                1..156942
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="19"
                /clone="RPCI-13_947L3"

ORIGIN
Query Match 84.0%; Score 16.8; DB 2; Length 156942;
Best Local Similarity 90.0%; Pred. No. 2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0,
1 TCGAAGCTGGGCAATCTGTG 20
|||||
124563 TCCAAAGTGGGCAAGTCTGTG 124582
db

```

Search completed: November 7, 2005, 06:41:59
Job time : 795.957 secs

This Page Blank (upside)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:04:21 ; Search time 230.638 Seconds
(without alignments)
513.336 Million cell updates/sec

Title: US-10-777-131A-3

Perfect score: 20

Sequence: 1 tcgaagtggtggcaactcgtg 20

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001s:*
5: geneseqn2001bs:*
6: geneseqn2002bs:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	ABZ79929	Abz79929 Mycobacter
2	20	100.0	1144	ABZ79953	Abz79953 Mycobacter
3	20	100.0	103765	AA199683_43	Continuation (44 o
4	20	100.0	110000	AA199682_43	Continuation (44 o
5	16	80.0	511	ADN13107	Adn13107 Human pro
6	16	80.0	2267	AA197922	Aa197922 Human neu
7	16	80.0	2267	AA198077	Aa198077 Human neu
8	15.8	79.0	602	AC119646	Ac119646 DNA clone
9	15.8	79.0	622	AC119645	Ac119645 DNA clone
10	15.8	79.0	663	ABZ53602	Abz53602 Human CDN
11	15.8	79.0	714	AC119647	Ac119647 DNA clone
12	15.8	79.0	837	ABZ53601	Abz53601 Human CDN
13	15.8	79.0	950	ABZ12053	Abz12053 Human pol
14	15.8	79.0	950	ADMA4571	Adma4571 Novel hum
15	15.8	79.0	1104	ADFS3637	Adfs3637 Murine hum
16	15.8	79.0	1165	AAZ23915	Aaz23915 Human dua
17	15.8	79.0	1165	ADN75961	Adn75961 Human sig
18	15.8	79.0	1165	ADN76019	Adn76019 Human DSP
19	15.8	79.0	1262	AAZ09495	Aaz09495 Human SGP
20	15.8	79.0	1422	AAZ29556	Aaz29556 M.catarth

21	15.8	79.0	1891	10	ADB63795	Adb63795 Human CDN
22	15.8	79.0	2241	5	AAS68586	Aas68586 DNA encod
23	15.8	79.0	2395	4	AB113174	Ab113174 Drosophill
24	15.8	79.0	3252	4	AB106532	Ab106532 Drosophill
25	15.8	79.0	3520	13	ADS49321	Ads49321 Bacterial
26	15.8	79.0	4596	4	AB102235	Ab102235 Drosophill
27	15.8	79.0	7149	4	AB102234	Ab102234 Drosophill
28	15.8	79.0	95001	12	ADH56439	Adh56439 Human hyp
29	15.8	79.0	110000	12	ADO78172_1	Continuation (2 of
30	15.8	79.0	117328	13	ABD32886	Abd32886 Mouse can
31	15.8	79.0	121162	3	AAC66548	Aac66548 Human kin
32	15.8	79.0	145606	12	ADP44617	Adp44617 Human alp
33	15.8	79.0	145980	6	ABQ81842	Abq81842 Bifidobac
34	15.8	79.0	349980	13	ADT05737	Adt05737 Haemophil
35	15.4	77.0	360	4	AA186669	Aa186669 Human pol
36	15.4	77.0	822	12	ADJ42411	Adj42411 Plant CDN
37	15.4	77.0	825	13	ADT45067	Adt45067 Bacterial
38	15.4	77.0	1191	11	ADM92655	Adm92655 SNP-conta
39	15.4	77.0	1928	2	AAQ32139	Aaq32139 Bauri pro
40	15.4	77.0	3337	4	AB114924	Ab114924 Drosophill
41	15.4	77.0	4343	13	ADR07326	Adr07326 Full leng
42	15.4	77.0	23075	5	ABA19112	Ab19112 Human ner
43	15.4	77.0	49939	9	ADB16928	Adb16928 Human DYX
44	15.4	77.0	110000	6	ABQ69245_09	Continuation (10 o
45	15.4	77.0	110000	6	ABQ67197_08	Continuation (9 of

ALIGNMENTS

RESULT 1	ABZ79929	standard; DNA; 20 BP.
ID	ABZ79929	
XX	ABZ79929;	
AC		
XX		
XX		
DT	19-MAY-2003	(first entry)
XX		
DE	Mycobacterium tuberculosis Rv3908-1-1 PCR primer SEQ ID NO:3.	
XX		
XX	Mycobacterium tuberculosis; mutT2; alka; ogt; Rv3908; mutY; Rv3909;	
KW	detection; multidrug resistance; multiple drug resistance; MDR;	
KM	infection; PCR primer; ss.	
XX		
OS	Mycobacterium tuberculosis.	
XX		
XX	Synthetic.	
PN	WO2003016562-A2.	
XX		
PD	27-FEB-2003.	
XX		
PF	14-AUG-2002; 2002WO-EP009679.	
XX		
PR	14-AUG-2001; 2001US-0311824P.	
PR	21-AUG-2001; 2001US-0313523P.	
XX		
PA	(INSP) INST PASTEUR.	
XX		
PI	Gicquel B;	
XX		
DR	WPI; 2003-256711/25.	
XX		
PT	Predicting the epidemic character of a Mycobacterium tuberculosis isolate	
XX	and/or the acquisition of multiple drug resistance (MDR) by the isolate	
PT	by detecting an alteration in the DNA repair system of the isolate.	
XX		
PS	Claim 32; Page 16; 83pp; English.	
XX		
CC	The present invention describes a method for predicting the epidemic	
CC	character of a Mycobacterium tuberculosis isolate and/or a selective	
CC	advantage to be maintained in the host and/or the acquisition of multiple	
CC	drug resistance (MDR) by the isolate comprising detecting an alteration	
CC	in the DNA repair system of the isolate. Also described: (1) detecting a	

CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a
CC polymnucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)
CC an Escherichia coli strain containing the plasmid pMYC2501; and (5)
CC detecting in a patient infected by Mycobacterium tuberculosis a higher
CC risk of being unable to eliminate the bacillus or of developing MDR
CC tuberculosis. The method is useful for predicting the epidemic character
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to
CC be maintained in the host and/or the acquisition of MDR by the isolate.
CC The present sequence represents a PCR primer for M. tuberculosis Rv3908,
CC which is used in the exemplification of the present invention
XX
SQ Sequence 20 BP; 5 A; 3 C; 8 G; 4 T; 0 U; 0 Other;

Query Match 100.0%; Score 20; DB 8; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGAAGTGGGCAATCGTG 20
Db 1 TCGAAGTGGGCAATCGTG 20
|||||
1 TCGAAGTGGGCAATCGTG 20

RESULT 2
ABZ79953
ID ABZ79953 standard; DNA; 1144 BP.
XX
AC ABZ79953;
XX
DT 19-MAY-2003 (first entry)
XX
DE Mycobacterium tuberculosis Rv3908 nucleotide sequence SEQ ID NO:20.
XX
KW Mycobacterium tuberculosis; mutT2; alkA; ogt; Rv3908; mutY; Rv3909;
KW detection; multidrug resistance; multiple drug resistance; MDR;
KW infection; gene; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO2003016562-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-EP009679.
XX
PR 14-AUG-2001; 2001US-0311824P.
PR 21-AUG-2001; 2001US-0313523P.
XX
PA (INSP) INST PASTEUR.
XX
PI Glcquel B;
XX
DR WPI; 2003-256711/25.
XX
PT Predicting the epidemic character of a Mycobacterium tuberculosis isolate
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate
PT by detecting an alteration in the DNA repair system of the isolate.
PS Disclosure; Fig 5A; 83pp; English.
XX
XX The present invention describes a method for predicting the epidemic
CC character of a Mycobacterium tuberculosis isolate and/or a selective
CC advantage to be maintained in the host and/or the acquisition of multiple
CC drug resistance (MDR) by the isolate comprising detecting an alteration
CC in the DNA repair system of the isolate. Also described: (1) detecting a
CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a
CC polymnucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)
CC an Escherichia coli strain containing the plasmid pMYC2501; and (5)
CC detecting in a patient infected by Mycobacterium tuberculosis a higher
CC risk of being unable to eliminate the bacillus or of developing MDR
CC tuberculosis. The method is useful for predicting the epidemic character
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to
CC be maintained in the host and/or the acquisition of MDR by the isolate.
CC The present sequence represents a M. tuberculosis Rv3908 nucleotide

CC sequence, which is used in the exemplification of the present invention
XX
SQ Sequence 1144 BP; 194 A; 405 C; 379 G; 166 T; 0 U; 0 Other;

Query Match 100.0%; Score 20; DB 8; Length 1144;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGAAGTGGGCAATCGTG 20
Db 7 TCGAAGTGGGCAATCGTG 26
|||||
7 TCGAAGTGGGCAATCGTG 26

RESULT 3
AA199683 43
Continuation (44 of 44) of AA199683 from base 4300001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AA199683 Accession AA199683
WP Fragment Name Begin End
WP AA199683_00 1 110000
WP AA199683_01 100001 210000
WP AA199683_02 200001 310000
WP AA199683_03 300001 410000
WP AA199683_04 400001 510000
WP AA199683_05 500001 610000
WP AA199683_06 600001 710000
WP AA199683_07 700001 810000
WP AA199683_08 800001 910000
WP AA199683_09 900001 1010000
WP AA199683_10 1000001 1110000
WP AA199683_11 1100001 1210000
WP AA199683_12 1200001 1310000
WP AA199683_13 1300001 1410000
WP AA199683_14 1400001 1510000
WP AA199683_15 1500001 1610000
WP AA199683_16 1600001 1710000
WP AA199683_17 1700001 1810000
WP AA199683_18 1800001 1910000
WP AA199683_19 1900001 2010000
WP AA199683_20 2000001 2110000
WP AA199683_21 2100001 2210000
WP AA199683_22 2200001 2310000
WP AA199683_23 2300001 2410000
WP AA199683_24 2400001 2510000
WP AA199683_25 2500001 2610000
WP AA199683_26 2600001 2710000
WP AA199683_27 2700001 2810000
WP AA199683_28 2800001 2910000
WP AA199683_29 2900001 3010000
WP AA199683_30 3000001 3110000
WP AA199683_31 3100001 3210000
WP AA199683_32 3200001 3310000
WP AA199683_33 3300001 3410000
WP AA199683_34 3400001 3510000
WP AA199683_35 3500001 3610000
WP AA199683_36 3600001 3710000
WP AA199683_37 3700001 3810000
WP AA199683_38 3800001 3910000
WP AA199683_39 3900001 4010000
WP AA199683_40 4000001 4110000
WP AA199683_41 4100001 4210000
WP AA199683_42 4200001 4310000
WP AA199683_43 4300001 4403765

Query Match 100.0%; Score 20; DB 4; Length 103765;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGAAGTGGGCAATCGTG 20
Db 85490 TCGAAGTGGGCAATCGTG 85509
|||||
1 TCGAAGTGGGCAATCGTG 20

RESULT 4


```
XX 13-SEP-2001.
PD 02-MAR-2001; 2001WO-JP001629.
XX 07-MAR-2000; 2000JP-00159195.
PR 07-MAR-2000; 2000JP-00159195.
XX (CHIB-) CHIBA PREFECTURE.
PA (HISM ) HISAMITSU PHARM CO LTD.
XX Nakagawara A;
PI WPI; 2001-565584/63.
XX Nucleic acids originating in gene expressed in human neuroblastoma,
PT useful as probe or primer in diagnosing prognosis of human neuroblastoma,
PT malignancy and susceptibility indicator or tumor marker for anti-cancer
PT agents.
XX Claim 1; Page 2945-2946; 2979pp; Japanese.
XX The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes
XX
XX Sequence 2267 BP; 755 A; 439 C; 483 G; 590 T; 0 U; 0 Other;
SQ
Query Match 80.0%; Score 16; DB 4; Length 2267;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 CGAAGTGGGCAATC 17
DB 1970 CGAAGTGGGCAATC 1985
RESULT 7
AAI98077
ID AAI98077 standard; cDNA; 2267 BP.
XX AAI98077;
XX 04-DEC-2001 (first entry)
XX Human neuroblastoma expressed polynucleotide SEQ ID NO 20.
DE Human neuroblastoma; ss.
XX Human; neuroblastoma; ss.
XX Homo sapiens.
XX WO200166733-A1.
XX 13-SEP-2001.
XX 02-MAR-2001; 2001WO-JP001631.
XX 07-MAR-2000; 2000JP-00159195.
PR 12-MAY-2000; 2000JP-00140387.
XX (CHIB-) CHIBA PREFECTURE.
PA (HISM ) HISAMITSU PHARM CO LTD.
XX Nakagawara A;
PI WPI; 2001-602630/68.
XX Nucleic acids for prognosis of human neuroblastoma comprise nucleic acids
PT expressed by human neuroblastomas.
PS Claim 1; Page 66-67; 159pp; Japanese.
```

```
XX The invention relates to nucleic acids (AAI98058-AAI98161) or their
CC homologues expressed by human neuroblastomas useful for detecting genes
CC expressed by neuroblastoma and for analysing their structure and
CC function. The nucleic acids are useful for the diagnosis and prognosis of
CC neuroblastoma
XX
XX Sequence 2267 BP; 755 A; 439 C; 483 G; 590 T; 0 U; 0 Other;
SQ
Query Match 80.0%; Score 16; DB 4; Length 2267;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 CGAAGTGGGCAATC 17
DB 1970 CGAAGTGGGCAATC 1985
RESULT 8
ACLI9646/C
ID ACLI9646 standard; DNA; 602 BP.
XX ACLI9646;
XX 17-OCT-2003 (first entry)
XX DNA clone originating in barley containing SNP encoding sequence #9637.
DE Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
XX gene; ss.
XX Hordeum vulgare.
XX WO2003057877-A1.
XX 17-JUL-2003.
XX 16-DEC-2002; 2002WO-IB005403.
XX 20-DEC-2001; 2001JP-00387059.
PR 20-DEC-2001; 2001JP-00387131.
PR 20-DEC-2001; 2001JP-00403289.
PR 20-DEC-2001; 2001JP-00403300.
PR 27-SEP-2002; 2002JP-00327515.
XX (UNNI-) UNIV JAPAN OKAYAMA.
XX Sato K, Takeda K, Kohara Y;
PI WPI; 2003-587127/55.
XX Single nucleotide polymorphism sites in barley varieties and DNA
PT sequences containing them for analysis and identification of barley
PT varieties and production of barley transformants with desired
PT characteristics.
XX Disclosure; SEQ ID XX; 284pp; Japanese.
XX The present invention relates to oligonucleotide clones originating in
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
CC varieties, identification of particular varieties and genotype-phenotype
CC analysis, isolation of specific genes and creation of new varieties by
CC transformation of barley varieties with them and production of new barley
CC varieties with desired properties. The present sequence represents an
CC oligonucleotide clone DNA sequence featured in the specification. The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published-pct-sequences
SQ Sequence 602 BP; 148 A; 145 C; 128 G; 181 T; 0 U; 0 Other;
Query Match 79.0%; Score 15.8; DB 9; Length 602;
```

Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGGTGGGCAATCGT 19
|||||
Db 530 TCGAAGGTGGGCAATCGT 512

RESULT 9

ACLI9645/C
ID ACLI9645 standard; DNA; 622 BP.

XX ACCLI9645;

DT 17-OCT-2003 (first entry)

DE DNA clone originating in barley containing SNP encoding sequence #9636.

KW Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
gene; ss.

XX Hordeum vulgare.

XX WO2003057877-A1.

PD 17-JUL-2003.

PF 16-DEC-2002; 2002WO-IB005403.

PR 20-DEC-2001; 2001JP-00387059.

PR 20-DEC-2001; 2001JP-00387131.

PR 20-DEC-2001; 2001JP-00403299.

PR 20-DEC-2001; 2001JP-00403300.

PR 27-SEP-2002; 2002JP-00327515.

XX (UYN1-) UNITV JAPAN OKAYAMA.

XX Sato K, Takeda K, Kohara Y;

XX WPI; 2003-587127/55.

PT Single nucleotide polymorphism sites in barley varieties and DNA
PT sequences containing them for analysis and identification of barley
PT varieties and production of barley transformants with desired
PT characteristics.

PS Disclosure; SEQ ID XX; 284bp; Japanese.

XX The present invention relates to oligonucleotide clones originating in
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
CC varieties, identification of particular varieties and genotype-phenotype
CC analysis, isolation of specific genes and creation of new varieties by
CC transformation of barley varieties with them and production of new barley
CC varieties with desired properties. The present sequence represents an
CC oligonucleotide clone DNA sequence featured in the specification. The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published-pct-sequences

XX Sequence 622 BP; 151 A; 154 C; 132 G; 185 T; 0 U; 0 Other;

Qy Query Match 79.0%; Score 15.8; DB 9; Length 622;

Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGGTGGGCAATCGT 19
|||||
Db 540 TCGAAGGTGGGCAATCGT 522

RESULT 10
ABSS53602/C

ID ABSS53602 standard; cDNA; 663 BP.

XX ABSS53602;

DT 29-NOV-2002 (first entry)

DE Human cDNA encoding Tyrosine specific protein phosphatase #2.

KW Human; ss; gene; tyrosine specific protein phosphatase; PTPase;
KW central nervous system disorder; Parkinson's disease; gout; diabetes;
KW Alzheimer's disease; dementia; motor neuron disease; multiple sclerosis;
KW Pick's disease; Creutzfeldt-Jacob dementia; schizophrenia; pain;
KW cognitive disorder; obesity; cachexia; anorexia; wasting disorders;
KW hypertension; stroke; osteoarthritis; breast cancer; reduced fertility;
KW pregnancy complication; depression.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 1..662

FT /tag= a
FT /product= "Tyrosine specific protein phosphatase"
FT /EC_number= "3.1.3.48"

XX WO200263014-A2.

XX 15-AUG-2002.

XX 04-FEB-2002; 2002WO-EP001099.

XX 05-FEB-2001; 2001US-0265872P.

XX (FARB) BAYER AG.

XX Koehler RH;

XX WPI; 2002-627563/67.

XX P-PsDB; ABG32808.

PT New isolated polynucleotide encoding a human tyrosine-specific protein
PT phosphatase (PTPase), useful for treating diseases associated with PTPase
PT dysfunction e.g. obesity or diabetes.

PS Claim 1; Fig 1; 109pp; English.

XX The invention relates to an isolated polynucleotide encoding a tyrosine -
CC specific protein phosphatase polypeptide (PTPase, appearing as ABSS53601
CC and ABSS53602), a polynucleotide which hybridizes under stringent
CC conditions to the PTPase nucleic acid, or a polynucleotide having a
CC sequence that deviates from PTPase nucleic acid due to the degeneration
CC of the genetic code. Also included are PTPase expression vectors,
CC transformed host cells, the purified PTPase, producing the PTPase,
CC detecting the PTPase or its nucleic acid, and screening for agents that
CC reduce or modulate the activity of PTPase. The expression vector or the
CC agent is useful for treating diseases associated with tyrosine-specific
CC protein phosphatase dysfunction e.g. a central nervous system disorder
CC (e.g. Parkinson's disease, Alzheimer's disease, dementia, motor neuron
CC disease, multiple sclerosis, Pick's disease, Creutzfeldt-Jacob dementia,
CC schizophrenia, pain and cognitive disorders), obesity cachexia, anorexia,
CC wasting disorders, diabetes, hypertension, stroke, gout, osteoarthritis,
CC breast cancer, reduced fertility, pregnancy complications and depression.
CC Many other diseases and disorders are listed in the specification. The
CC present sequence encodes a human PTPase

XX Sequence 663 BP; 146 A; 199 C; 214 G; 104 T; 0 U; 0 Other;

Qy Query Match 79.0%; Score 15.8; DB 6; Length 663;

Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGGTGGGCAATCGT 19
|||||
Db 359 TCGAAGGTGGGCAATCGT 341

```
RESULT 11
ACLI9647/C
ID ACLI9647 standard; DNA; 714 BP.
AC
ACLI9647;
XX
DT 17-OCT-2003 (first entry)
XX
DE DNA clone originating in barley containing SNP encoding sequence #9638.
XX
KW Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KW gene; ss.
XX
OS Hordeum vulgare.
XX
PN WO2003057877-A1.
XX
PD 17-JUL-2003.
XX
PE 16-DEC-2002; 2002MO-1B005403.
XX
PR 20-DEC-2001; 2001JP-00387059.
PR 20-DEC-2001; 2001JP-00387131.
PR 20-DEC-2001; 2001JP-00403299.
PR 20-DEC-2001; 2001JP-00403300.
PR 27-SEP-2002; 2002JP-00327515.
XX
PA (UYN1-) UNIV JAPAN OKAYAMA.
XX
PI Sato K, Takeda K, Kohara Y;
XX
DR WPI; 2003-587127/55.
XX
PT Single nucleotide polymorphism sites in barley varieties and DNA
PT sequences containing them for analysis and identification of barley
PT varieties and production of barley transformants with desired
PT characteristics.
XX
PS Disclosure; SEQ ID XX; 284bp; Japanese.
XX
CC The present invention relates to oligonucleotide clones originating in
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
CC varieties, identification of particular varieties and genotype-phenotype
CC analysis, isolation of specific genes and creation of new varieties by
CC transformation of barley varieties with them and production of new barley
CC varieties with desired properties. The present sequence represents an
CC oligonucleotide clone DNA sequence featured in the specification. The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published-pct-sequences
XX
SQ Sequence 714 BP; 190 A; 165 C; 150 G; 207 T; 0 U; 2 Other;
XX
Query Match 79.0%; Score 15.8; DB 9; Length 714;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TCGAAGGTGGGCAATCGT 19
DB 688 TCGAAGGTGGGCAATCGT 670
XX
RESULT 12
ABSS3601/C
ID ABSS3601 standard; cDNA; 837 BP.
AC
ABSS3601;
XX
DT 29-NOV-2002 (first entry)
XX
```

```
DE Human cDNA encoding Tyrosine specific protein phosphatase #1.
XX
KW Human; ss; gene; tyrosine specific protein phosphatase; PTPase;
KW central nervous system disorder; Parkinson's disease; gout; diabetes;
KW Alzheimer's disease; dementia; motor neuron disease; multiple sclerosis;
KW Pick's disease; Creutzfeldt-Jacob dementia; schizophrenia; pain;
KW cognitive disorder; obesity; cachexia; anorexia; wasting disorder;
KW hypertension; stroke; osteoarthritis; breast cancer; reduced fertility;
KW pregnancy complication; depression.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 46..837
FT /tag=a
FT /product="Tyrosine specific protein phosphatase"
FT /EC_number="3.1.3.48"
XX
PN WO200263014-A2.
XX
PD 15-AUG-2002.
XX
PE 04-FEB-2002; 2002MO-EP001099.
XX
PR 05-FEB-2001; 2001US-0265872P.
XX
PA (FARB ) BAYER AG.
XX
PI Koehler RH;
XX
DR WPI; 2002-627563/67.
DR P-PSDB; ABG32807.
XX
XX
PT New isolated polynucleotide encoding a human tyrosine-specific protein
PT phosphatase (PTPase), useful for treating diseases associated with PTPase
PT dysfunction e.g. obesity or diabetes.
XX
PS Claim 1; Fig 5; 109pp; English.
XX
CC The invention relates to an isolated polynucleotide encoding a tyrosine
CC specific protein phosphatase polypeptide (PTPase, appearing as ABSS3601
CC and ABSS3602), a polynucleotide which hybridizes under stringent
CC conditions to the PTPase nucleic acid, or a polynucleotide having a
CC sequence that deviates from PTPase nucleic acid due to the degeneration
CC of the genetic code. Also included are PTPase expression vectors,
CC transformed host cells, the purified PTPase, producing the PTPase,
CC detecting the PTPase or its nucleic acid, and screening for agents that
CC reduce or modulate the activity of PTPase. The expression vector or the
CC agent is useful for treating diseases associated with tyrosine-specific
CC protein phosphatase dysfunction e.g. a central nervous system disorder
CC (e.g. Parkinson's disease, Alzheimer's disease, dementia, motor neuron
CC disease, multiple sclerosis, Pick's disease, Creutzfeldt-Jacob dementia,
CC schizophrenia, pain and cognitive disorders), obesity cachexia, anorexia,
CC wasting disorders, diabetes, hypertension, stroke, gout, osteoarthritis,
CC breast cancer, reduced fertility, pregnancy complications and depression.
CC Many other diseases and disorders are listed in the specification. The
CC present sequence encodes a human PTPase
XX
SQ Sequence 837 BP; 203 A; 231 C; 252 G; 151 T; 0 U; 0 Other;
XX
Query Match 79.0%; Score 15.8; DB 6; Length 837;
Best Local Similarity 89.5%; Pred. No. 3.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TCGAAGGTGGGCAATCGT 19
DB 533 TCGAAGGTGGGCAATCGT 515
XX
RESULT 13
ABZ12053/C
ID ABZ12053 standard; cDNA; 950 BP.
AC
ABZ12053;
XX
DT
XX
```



```

PD 06-NOV-2003.
XX
XX 23-APR-2003; 2003WO-JP005175.
XX
XX 23-APR-2002; 2002JP-00120709.
XX
XX 04-DEC-2002; 2002JP-00352308.
XX
XX (RIKE ) RIKEN KK.
XX (DNAF-) DNAFORM KK.
XX (MITU ) MITSUBISHI CHEM CORP.
XX
XX Hayashizaki Y, Kamiya M, Kubodera H;
XX
XX WPI; 2003-854398/79.
XX
XX P-PSDB; ADF53644.
XX
XX Proteins with phosphatase activity useful for screening for substances
XX that affect their activity.
XX
XX Claim 4; SEQ ID NO 5; 170pp; Japanese.
XX
XX The invention relates to novel isolated proteins comprising an amino acid
XX sequence represented by any of SEQ ID Nos: 8 to 14, 25 and 26; and a
XX protein comprising an amino acid sequence derived from an amino acid
XX sequence represented by any of SEQ ID Nos: 8 to 14, 25 and 26 by
XX deletion, substitution and/or addition of one to several amino acids and
XX having a phosphatase activity. The invention further comprises: the DNA
XX of optionally mutated sequences SEQ ID Nos: 1-7, 23 and 24 or those that
XX hybridise to them; DNA and cDNA encoding the proteins; vectors containing
XX the DNA; cells containing the vectors; proteins produced by the cells;
XX whole or fragments of antibodies against the proteins; screening for
XX substances that affect the expression and activity of the proteins; a
XX computer readable recording medium; and a solid support bound to the DNA.
XX The novel phosphatase proteins are useful for screening for substances
XX that affect the activity of the proteins, which have potential use in the
XX development of drugs for treating associated disorders. This
XX polynucleotide represents the DNA of an optionally mutated sequence of
XX the invention.
XX
XX SQ Sequence 1104 BP; 255 A; 287 C; 333 G; 229 T; 0 U; 0 Other;
XX
XX Query Match 79.0%; Score 15.8; DB 10; Length 1104;
XX Best Local Similarity 89.5%; Pred. No. 3.8e+02;
XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 TCGAAGTGGGCAATCGT 19
XX |||||
XX 453 TCGAAGTGGGCAATCGT 435
XX

```

Search completed: November 7, 2005, 05:40:19
 Job time : 237.638 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 19:19:37 ; Search time 67.8723 Seconds
(without alignments)
482.163 Million cell updates/sec

Title: US-10-777-131A-3

Perfect score: 20

Sequence: 1 tcgaagtcggcgcaatcgtg 20

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405566

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing filter 45 summaries

Database :

Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6C COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	4403765	3 US-09-103-840A-2	Sequence 2, Appl1
2	20	100.0	4411529	3 US-09-103-840A-1	Sequence 1, Appl1
3	15.8	79.0	601	4 US-09-949-016-117629	Sequence 117629,
4	15.8	79.0	601	4 US-09-949-016-117655	Sequence 141204,
5	15.8	79.0	601	4 US-09-949-016-141204	Sequence 935, App
6	15.8	79.0	950	4 US-09-799-451-935	Sequence 971, App
7	15.8	79.0	1422	4 US-09-701-896-1	Sequence 16671, A
8	15.8	79.0	3615	3 US-09-221-017B-971	Sequence 15741, A
9	15.8	79.0	16112	4 US-09-949-016-15671	Sequence 12006, A
10	15.8	79.0	18264	4 US-09-949-016-15671	Sequence 14999, A
11	15.8	79.0	33753	4 US-09-949-016-15741	Sequence 15000, A
12	15.8	79.0	33753	4 US-09-949-016-15741	Sequence 14826, A
13	15.8	79.0	46288	4 US-09-949-016-14999	Sequence 13487, A
14	15.8	79.0	46311	4 US-09-949-016-15000	Sequence 15274, A
15	15.8	79.0	113876	4 US-09-949-016-14828	Sequence 14690, A
16	15.8	79.0	113876	4 US-09-949-016-14829	Sequence 14691, A
17	15.8	79.0	113876	4 US-09-949-016-14829	Sequence 14692, A
18	15.8	79.0	113876	4 US-09-949-016-14829	Sequence 14693, A
19	15.8	79.0	113876	4 US-09-949-016-14829	Sequence 14694, A
20	15.4	77.0	112042	4 US-09-949-016-14821	Sequence 14695, A
21	15.4	77.0	12797	4 US-09-949-016-15274	Sequence 14696, A
22	15.4	77.0	15499	4 US-09-949-016-14690	Sequence 14697, A
23	15.4	77.0	15499	4 US-09-949-016-14691	Sequence 14698, A
24	15.4	77.0	15499	4 US-09-949-016-14692	Sequence 14699, A
25	15.4	77.0	44789	4 US-09-949-016-13909	Sequence 12523, A
26	15.4	77.0	45693	4 US-09-949-016-13523	Sequence 12524, A
27	15.4	77.0	45693	4 US-09-949-016-12678	Sequence 12679, A

C 28	15.4	77.0	45693	4	US-09-949-016-12679	Sequence 12679, A
C 29	15.4	77.0	50269	4	US-09-949-016-17598	Sequence 17598, A
C 30	15.2	76.0	365	4	US-09-270-767-28333	Sequence 28333, A
C 31	15.2	76.0	773	4	US-09-270-767-13748	Sequence 13748, A
C 32	15.2	76.0	780	4	US-09-724-623-156	Sequence 56, Appl
C 33	15.2	76.0	846	4	US-09-583-110-2248	Sequence 2248, Ap
C 34	15.2	76.0	849	4	US-09-107-433-470	Sequence 470, App
C 35	15.2	76.0	934	4	US-09-270-767-12547	Sequence 12547, A
C 36	15.2	76.0	1542	3	US-09-385-028-16	Sequence 16, Appl
C 37	15.2	76.0	1542	4	US-09-726-614-16	Sequence 16, Appl
C 38	15.2	76.0	1542	4	US-09-385-040-16	Sequence 16, Appl
C 39	15.2	76.0	11604	3	US-09-385-028-13	Sequence 13, Appl
C 40	15.2	76.0	11604	4	US-09-726-614-13	Sequence 13, Appl
C 41	15.2	76.0	11604	4	US-09-385-028-13	Sequence 13, Appl
C 42	15.2	76.0	15079	3	US-09-385-028-1	Sequence 1, Appl
C 43	15.2	76.0	15079	4	US-09-726-614-1	Sequence 1, Appl
C 44	15.2	76.0	15120	4	US-09-385-040-1	Sequence 1, Appl
C 45	15.2	76.0	36412	4	US-08-311-731A-132	Sequence 132, App

ALIGNMENTS

```
RESULT 1
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2
Query Match          100.0%; Score 20; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. NO. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TCGAAGTGGCGCAATCTG 20
DB 4385490 TCGAAGTGGCGCAATCTG 4385509
RESULT 2
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
```

SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 441529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 100.0%; Score 20; DB 3; Length 441529;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGAAGTGGCGCAATCGTG 20
DB 4393252 TCGAAGTGGCGCAATCGTG 4393271

RESULT 3
US-09-949-016-117629/c
Sequence 117629, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 117629
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-117629

Query Match 79.0%; Score 15.8; DB 4; Length 601;
Best Local Similarity 89.5%; Pred. No. 93;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 CGAAGTGGCGCAATCGTG 20
DB 322 CCAAGTGGCGCAATCATG 304

RESULT 4
US-09-949-016-117655/c
Sequence 117655, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 117655
LENGTH: 601
TYPE: DNA

ORGANISM: Human
US-09-949-016-117655

Query Match 79.0%; Score 15.8; DB 4; Length 601;
Best Local Similarity 89.5%; Pred. No. 93;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 CGAAGTGGCGCAATCGTG 20
DB 322 CCAAGTGGCGCAATCATG 304

RESULT 5
US-09-949-016-141204
Sequence 141204, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 141204
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-141204

Query Match 79.0%; Score 15.8; DB 4; Length 601;
Best Local Similarity 89.5%; Pred. No. 93;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 CGAAGTGGCGCAATCGTG 20
DB 239 CCAAGTGGCGCAGATCGTG 257

RESULT 6
US-09-799-451-935/c
Sequence 935, Application US/09799451
Patent No. 6783969
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhou, Ping
APPLICANT: Goodrich, Ryle
APPLICANT: Asundi, Vinod
APPLICANT: Ren, Feiyun
APPLICANT: Zhang, Jie
APPLICANT: Xue, Aiding J.
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Ma, Yundong
APPLICANT: Yamazaki, Victoria
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Zhiwei
APPLICANT: Wang, Yonghong
APPLICANT: Yang, Dunrui
APPLICANT: Wenman, Tom
APPLICANT: Ghosh, Reena
APPLICANT: Drmanac, Radote T.
TITLE OF INVENTION: No. 6783969el Nucleic Acids and
Polypeptides
FILE REFERENCE: 803
CURRENT APPLICATION NUMBER: US/09/799,451

```

; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 948
; SOFTWARE: PE_FL_genes Version 2.0
; SEQ ID NO 935
; LENGTH: 950
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (287)..(946)
; US-09-799-451-935

Query Match          79.0%; Score 15.8; DB 4; Length 950;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 TCGAAGTGGGCAATCGT 19
        |||||
Db      645 TCGAAGTGGGCAATCGT 627

RESULT 7
US-09-701-896-1
; Sequence 1, Application US/09701896
; Patent No. 6706269
; GENERAL INFORMATION:
; APPLICANT: Jean-Louise Ruelie
; APPLICANT: Carlotia Vimalis y de Basozola
; TITLE OF INVENTION: No. 6706269el Compounds
; FILE REFERENCE: BM45325
; CURRENT APPLICATION NUMBER: US/09/701,896
; PRIOR FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: 9812163.5
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1422
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
; US-09-701-896-1

Query Match          79.0%; Score 15.8; DB 4; Length 1422;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 TCGAAGTGGGCAATCGT 19
        |||||
Db      1244 TCGAAGTGGGCAATCGT 1262

RESULT 8
US-09-221-017B-971/c
; Sequence 971, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: ROSE, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
```

```

; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Montoy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 971:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURES:
; NAME/KEY: misc_feature
; LOCATION: 1...3615
; US-09-221-017B-971

Query Match          79.0%; Score 15.8; DB 3; Length 3615;
Best Local Similarity 89.5%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 TCGAAGTGGGCAATCGT 19
        |||||
Db      421 TCGAATGTGGGCAATCT 403

RESULT 9
US-09-949-016-16671/c
; Sequence 16671, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16671
; LENGTH: 16112
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-16671

Query Match          79.0%; Score 15.8; DB 4; Length 16112;
```

Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGAAGTGGGCAATCGTG 20

Db 1019 CCAAGTGGGCAATCATG 1001

RESULT 10

US-09-949-016-16469/C
; Sequence 16469, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16469
; LENGTH: 18264
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16469

Query Match 79.0%; Score 15.8; DB 4; Length 18264;
Best Local Similarity 89.5%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGAAGTGGGCAATCGTG 20

Db 18003 CCAAGTGGGCAATCATG 17985

RESULT 11
US-09-949-016-15741/C

; Sequence 15741, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15741
; LENGTH: 33753
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15741

Query Match 79.0%; Score 15.8; DB 4; Length 33753;
Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGAAGTGGGCAATCGTG 20

Db 21336 CCAAGTGGGCAATCGTG 21318

RESULT 12

US-09-949-016-12006/C
; Sequence 12006, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12006
; LENGTH: 33756
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12006

Query Match 79.0%; Score 15.8; DB 4; Length 33756;
Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGAAGTGGGCAATCGTG 20

Db 21340 CCAAGTGGGCAATCATG 21322

RESULT 13
US-09-949-016-14999

; Sequence 14999, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14999
; LENGTH: 46288
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14999

Query Match 79.0%; Score 15.8; DB 4; Length 46288;
Best Local Similarity 89.5%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGAAGTGGGCAATCGTG 20

Db 10787 CCAAGTGGGCAATCATG 10805

RESULT 14
US-09-949-016-15000

```
; Sequence 15000, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15000
; LENGTH: 46311
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15000

Query Match          79.0%; Score 15.8; DB 4; Length 46311;
Best Local Similarity 89.5%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      2 CGAAGTGGCAATCGTG 20
Db      10810 CCAAGTGGCAATCATG 10828

RESULT 15
US-09-949-016-14828
; Sequence 14828, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14828
; LENGTH: 113876
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14828

Query Match          79.0%; Score 15.8; DB 4; Length 113876;
Best Local Similarity 89.5%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      2 CGAAGTGGCAATCGTG 20
Db      79450 CCAAGTGGCAATCATG 79468

Search completed: November 6, 2005, 23:10:27
Job time : 85.8723 secs
```

This page blank (spec)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 20:56:42 ; Search time 498.723 Seconds
(without alignments)
331.640 Million cell updates/sec

Title: US-10-777-131a-3

Perfect score: 20

Sequence: 1 tcgaagctggcgcaatcgtg 20

Scoring table:

IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 9794790 seqs, 413490567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
25: /cgn2_6/ptodata/2/pubpna/US11_PUBCOMB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
27: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
28: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	16	US-10-216-817-3
2	20	100.0	20	22	US-10-777-131a-3
3	20	100.0	1144	16	US-10-216-817-20
4	20	100.0	1144	22	US-10-777-131a-20
5	17.4	87.0	937	20	US-10-437-963-37441

c	6	16.8	84.0	241	21	US-10-425-115-44578	Sequence 44578, A
	7	16.8	84.0	2589	21	US-10-425-115-175890	Sequence 175890, A
	8	16.8	84.0	5688	21	US-10-425-115-175894	Sequence 175894, A
	9	16.4	82.0	588	13	US-09-925-065A-830421	Sequence 830421, A
c	10	16.4	82.0	1065	19	US-10-424-599-105971	Sequence 105971, A
	11	16	80.0	2267	18	US-10-220-891-20	Sequence 20, Appl
	12	15.8	79.0	484	13	US-09-925-065A-140048	Sequence 140048, A
	13	15.8	79.0	600	24	US-10-972-079-52425	Sequence 52425, A
	14	15.8	79.0	600	26	US-11-060-756-3208	Sequence 3208, Ap
	15	15.8	79.0	600	26	US-11-060-756-7480	Sequence 7480, Ap
	16	15.8	79.0	611	13	US-09-925-065A-424836	Sequence 424836, A
	17	15.8	79.0	653	13	US-09-925-065A-918948	Sequence 918948, A
	18	15.8	79.0	658	13	US-09-925-065A-918085	Sequence 918085, A
	19	15.8	79.0	661	20	US-10-767-701-26185	Sequence 26185, A
c	20	15.8	79.0	663	19	US-10-470-992-4	Sequence 4, Appl1
	21	15.8	79.0	664	14	US-10-027-632-105474	Sequence 105474, A
	22	15.8	79.0	664	14	US-10-027-632-105475	Sequence 105475, A
	23	15.8	79.0	664	18	US-10-027-632-105474	Sequence 105474, A
	24	15.8	79.0	664	18	US-10-027-632-105475	Sequence 105475, A
c	25	15.8	79.0	729	14	US-10-027-632-16570	Sequence 16570, A
	26	15.8	79.0	729	14	US-10-027-632-16571	Sequence 16571, A
	27	15.8	79.0	729	18	US-10-027-632-16570	Sequence 16570, A
	28	15.8	79.0	729	18	US-10-027-632-16571	Sequence 16571, A
c	29	15.8	79.0	800	14	US-10-027-632-168257	Sequence 168257, A
	30	15.8	79.0	800	14	US-10-027-632-168258	Sequence 168258, A
c	31	15.8	79.0	800	18	US-10-027-632-168257	Sequence 168257, A
	32	15.8	79.0	800	18	US-10-027-632-168258	Sequence 168258, A
c	33	15.8	79.0	837	19	US-10-470-992-1	Sequence 1, Appl1
	34	15.8	79.0	905	14	US-10-027-632-9777	Sequence 9777, Ap
	35	15.8	79.0	905	18	US-10-027-632-9777	Sequence 9777, Ap
c	36	15.8	79.0	950	19	US-10-302-172-935	Sequence 935, App
	37	15.8	79.0	1165	9	US-09-847-519A-1	Sequence 1, Appl1
c	38	15.8	79.0	1165	19	US-10-444-795B-786	Sequence 786, App
	39	15.8	79.0	1262	19	US-10-168-506-5	Sequence 5, Appl1
c	40	15.8	79.0	1262	22	US-10-838-181-5	Sequence 5, Appl1
	41	15.8	79.0	1891	18	US-10-104-047-1949	Sequence 1949, Ap
c	42	15.8	79.0	2070	21	US-10-425-115-113355	Sequence 113355, A
c	43	15.8	79.0	2241	24	US-10-450-763-4390	Sequence 4390, Ap
	44	15.8	79.0	2263	21	US-10-425-115-153796	Sequence 153796, A
	45	15.8	79.0	2395	26	US-11-097-143-17002	Sequence 17002, A

ALIGNMENTS

RESULT 1

US-10-216-817-3

Sequence 3, Application US/10216817

Publication No. US20030129619A1

GENERAL INFORMATION:

APPLICANT: GICQUEL, BRIGITTE

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG

TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS

TITLE OF INVENTION: IN GENES OF THE MULT FAMILY

FILE REFERENCE: 03495.0233-00000

CURRENT APPLICATION NUMBER: US/10/216,817

CURRENT FILING DATE: 2002-11-13

PRIOR FILING DATE: 2001-08-14

PRIOR APPLICATION NUMBER: 60/311,824

PRIOR FILING DATE: 2001-08-21

PRIOR APPLICATION NUMBER: 60/313,523

NUMBER OF SEQ ID NOS: 32

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 3

LENGTH: 20

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Primer

US-10-216-817-3

Query Match 100.0%; Score 20; DB 16; Length 20;

Best Local Similarity 100.0%; Pred. No. 5.7;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGAAGTGGGCAATCGTG 20
Db 1 TCGAAGTGGGCAATCGTG 20

RESULT 2

US-10-777-131A-3
; Sequence 3, Application US/10777131A
; Publication No. US20050026216A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; TITLE OF INVENTION: IN GENES OF THE mltcT FAMILY
; FILE REFERENCE: 02356.0090-00000
; CURRENT FILING DATE: 2004-02-13
; PRIOR APPLICATION NUMBER: PCT/EP02/09679
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 3
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
US-10-777-131A-3

Query Match 100.0%; Score 20; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGAAGTGGGCAATCGTG 20
Db 1 TCGAAGTGGGCAATCGTG 20

RESULT 3

US-10-216-817-20
; Sequence 20, Application US/10216817
; Publication No. US20030129619A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; TITLE OF INVENTION: IN GENES OF THE mltcT FAMILY
; FILE REFERENCE: 03495.0233-00000
; CURRENT FILING DATE: US/10/216,817
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-14
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 1144
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-216-817-20

Query Match 100.0%; Score 20; DB 16; Length 1144;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGAAGTGGGCAATCGTG 20

Db 7 TCGAAGTGGGCAATCGTG 26

RESULT 4

US-10-777-131A-20
; Sequence 20, Application US/10777131A
; Publication No. US20050026216A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; TITLE OF INVENTION: IN GENES OF THE mltcT FAMILY
; FILE REFERENCE: 02356.0090-00000
; CURRENT FILING DATE: 2004-02-13
; PRIOR APPLICATION NUMBER: PCT/EP02/09679
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 20
; LENGTH: 1144
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-777-131A-20

Query Match 100.0%; Score 20; DB 22; Length 1144;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGAAGTGGGCAATCGTG 20
Db 7 TCGAAGTGGGCAATCGTG 26

RESULT 5

US-10-437-963-37441/C
; Sequence 37441, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbasuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT FILING DATE: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 37441
; LENGTH: 937
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_41171C.1
US-10-437-963-37441

Query Match 87.0%; Score 17.4; DB 20; Length 937;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGAAGTGGGCAATCGT 19
Db 758 TCGAAGTGGGCAATCAT 740


```
RESULT 6
US-10-425-115-44578/c
; Sequence 44578, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT FILING DATE: 2003-04-28
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 44578
; LENGTH: 241
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_14065C.1
US-10-425-115-44578

Query Match      84.0%; Score 16.8; DB 21; Length 241;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGAAGTGGGCAATCGTG 20
    |||||
DB 99 TTGAAGTGGGAATTCGTG 80

RESULT 7
US-10-425-115-175990
; Sequence 175990, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT FILING DATE: 2003-04-28
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 175990
; LENGTH: 2589
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) (2589)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_92098C.1
US-10-425-115-175990

Query Match      84.0%; Score 16.8; DB 21; Length 2589;
Best Local Similarity 90.0%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGAAGTGGGCAATCGTG 20
    |||||
DB 1372 TCGAAGTGGTCATATCGTG 1391

RESULT 8
US-10-425-115-175994
; Sequence 175994, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT FILING DATE: 2003-04-28
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 175994
; LENGTH: 5658
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_92100C.1
US-10-425-115-175994

Query Match      84.0%; Score 16.8; DB 21; Length 5658;
Best Local Similarity 90.0%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGAAGTGGGCAATCGTG 20
    |||||
DB 1558 TCGAAGTGGTCATATCGTG 1577

RESULT 9
US-09-925-065A-830421
; Sequence 830421, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; CURRENT APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 830421
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION: Clone ID: MRT4577_92100C.1
US-09-925-065A-830421

Query Match      82.0%; Score 16.4; DB 13; Length 588;
Best Local Similarity 94.4%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GAAAGTGGGCAATCGTG 20
    |||||
DB 144 GAAAGTGGGCAATCGTG 161

RESULT 10
US-10-424-599-105971/c
; Sequence 105971, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
```

```

; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 105971
; LENGTH: 1065
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_66708C.1
US-10-424-599-105971

Query Match      82.0%; Score 16.4; DB 19; Length 1065;
Best Local Similarity 94.4%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3  CGAAGTGGGCAATCGTG 20
DB      857  GAAAGTGGGCAATAGTG 840

RESULT 11
US-10-220-891-20
; Sequence 20, Application US/10220891
; Publication No. US20030207286A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWARA, AKIRA
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES HAVING CHARACTERISTICS OF ENHANCED
; TITLE OF INVENTION: EXPRESSION IN HUMAN NEUROBLASTOMA WITH FAVORABLE PROGNOSIS
; TITLE OF INVENTION: BASED ON COMPARISON BETWEEN HUMAN NEUROBLASTOMA WITH FAVORABLE
; TITLE OF INVENTION: PROGNOSIS AND HUMAN NEUROBLASTOMA WITH UNFAVORABLE PROGNOSIS
; FILE REFERENCE: 7388-73435
; CURRENT APPLICATION NUMBER: US/10/220,891
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: JP 2000/140387
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: JP 2000/159195
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 2267
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-220-891-20

Query Match      80.0%; Score 16; DB 18; Length 2267;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2  CGAAGTGGGCAATC 17
DB      1970  CGAAGTGGGCAATC 1985

RESULT 12
US-09-925-065A-140048
; Sequence 140048, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
```

```

; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140048
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-140048

Query Match      79.0%; Score 15.8; DB 13; Length 484;
Best Local Similarity 89.5%; Pred. No. 7.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2  CGAAGTGGGCAATCGTG 20
DB      123  CGAAGTGGGCAATCATG 141

RESULT 13
US-10-972-079-52425/c
; Sequence 52425, Application US/10972079
; Publication No. US2005015317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF
; TITLE OF INVENTION: LIVESTOCK
; FILE REFERENCE: MM11110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52425
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894303000_4
US-10-972-079-52425

Query Match      79.0%; Score 15.8; DB 24; Length 600;
Best Local Similarity 89.5%; Pred. No. 7.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2  CGAAGTGGGCAATCGTG 20
DB      500  CGAAGTGGGCAAAACCTG 482

RESULT 14
US-11-060-756-3208/c
; Sequence 3208, Application US/11060756
; Publication No. US20050221354A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042000)
; CURRENT APPLICATION NUMBER: US/11/060,756
; CURRENT FILING DATE: 2005-02-18
; NUMBER OF SEQ ID NOS: 303284
```

```

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3208
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-060-756-3208

```

```

Query Match          79.0%; Score 15.8; DB 26; Length 600;
Best Local Similarity 89.5%; Pred. No. 7.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 TCGAAGGTGGGCAATCGT 19
          |||||
Db      161 TCGAAGGTGGGCAATCGT 143

```

```

RESULT 15
US-11-060-756-7480/c
; Sequence 7480, Application US/11060756
; Publication No. US2005021354A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101093 (031896-042000)
; CURRENT APPLICATION NUMBER: US/11/060,756
; CURRENT FILING DATE: 2005-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7480
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-060-756-7480

```

```

Query Match          79.0%; Score 15.8; DB 26; Length 600;
Best Local Similarity 89.5%; Pred. No. 7.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 TCGAAGGTGGGCAATCGT 19
          |||||
Db      161 TCGAAGGTGGGCAATCGT 143

```

Search completed: November 7, 2005, 05:21:11
Job time : 501.723 secs

THE DOG PARK (1994)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 02:38:52 ; Search time 2000.85 Seconds
(without alignments)
380.481 Million cell updates/sec

Title: US-10-777-131A-3

Perfect score: 20

Sequence: 1 tcgaagtggtggcgaatcgtg 20

Scoring table: IDENTITY_NUC

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Listing first 45 summaries

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_est4: *
5: gb_est5: *
6: gb_est6: *
7: gb_est7: *
8: gb_est8: *
9: gb_est9: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18.4	92.0	592	7	CV251860 WS0118.B2
C 2	17.4	87.0	633	7	AZ564202 RPT-23-2
C 3	17	85.0	559	7	CK124820 BES182411
C 4	17	85.0	790	2	BF627004 HVSMB000
C 5	16.8	84.0	391	9	CG049330 PU1FO4TD
C 6	16.8	84.0	452	9	CL219748 ZM8B044
C 7	16.8	84.0	511	9	CC803612 ih17b02.9
C 8	16.8	84.0	548	9	BU098971 BU098971
C 9	16.8	84.0	553	5	BQ097494 ph03d06.y
C 10	16.8	84.0	607	8	AZ471085 IM0285D02
C 11	16.8	84.0	608	4	BJ460007 BJ460007
C 12	16.8	84.0	627	1	AV910592 AV910592
C 13	16.8	84.0	642	4	BJ462359 BJ462359
C 14	16.8	84.0	646	4	BJ452459 BJ452459
C 15	16.8	84.0	658	1	AV920519 AV920519
C 16	16.8	84.0	674	8	AZ574524 329PFC09
C 17	16.8	84.0	684	4	BJ460029 BJ460029
C 18	16.8	84.0	681	7	CV232798 WS0198.B2
C 19	16.8	84.0	685	1	AV919237 AV919237
C 20	16.8	84.0	701	1	AV914195 AV914195
C 21	16.8	84.0	702	1	BU448654 BU448654
C 22	16.8	84.0	711	9	CC671727 OGUPP72TV
C 23	16.8	84.0	728	4	BJ462046 BJ462046
C 24	16.8	84.0	803	8	BH574525 BOCJM13TF

25	16.8	84.0	827	9	CV538423	CV538423 OP_Ba001
26	16.8	84.0	828	9	CG004578	CG004578 ZUAAS80TH
27	16.8	84.0	843	8	BZ701269	BZ701269 PUCR48TD
28	16.8	84.0	848	4	BI116452	BI116452 602868878
C 29	16.8	84.0	941	4	BF965968	BF965968 602277143
C 30	16.8	84.0	947	4	CC412703	CC412703 PUHNM18TB
C 31	16.8	84.0	956	8	CC412704	CC412704 PUHNM18TD
C 32	16.8	84.0	1035	7	CL451811	CL451811 ZM8B047
C 33	16.8	84.0	1041	7	CF581680	CF581680 AGENCOURT
C 34	16.8	84.0	1172	8	CC268595	CC268595 CH261-67F
C 35	16.8	84.0	1604	2	BE613467	BE613467 601504233
C 36	16.4	82.0	328	1	AA861299	AA861299 ak33h05.x
C 37	16.4	82.0	329	1	AT425088	AT425088 t93hb04.x
C 38	16.4	82.0	360	1	AT221296	AT221296 g974f01.x
C 39	16.4	82.0	417	5	BU831776	BU831776 T025F06.P
C 40	16.4	82.0	443	5	BU821612	BU821612 UB25CP808
C 41	16.4	82.0	484	9	CL876478	CL876478 abf13a11.
C 42	16.4	82.0	490	1	AT161906	AT161906 A009P37U
C 43	16.4	82.0	498	7	CF230920	CF230920 PtAC0014H
C 44	16.4	82.0	511	1	AV699630	AV699630 AV699630
C 45	16.4	82.0	526	9	AG246983	AG246983 Lotus cor

ALIGNMENTS

RESULT 1
LOCUS CV251860/c 592 bp mRNA linear EST 22-SEP-2004
DEFINITION WS0118.B21_P22_PT-P-FL-A-2 Populus balsamifera subsp. trichocarpa
CDNA clone WS0118_P22 3', mRNA sequence.

ACCESSION CV251860
VERSION CV251860.1 GI:52504835
KEYWORDS
SOURCE
ORGANISM

Populus balsamifera subsp. trichocarpa (Populus trichocarpa)
Populus balsamifera subsp. trichocarpa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Malpighiales; Salicaceae; Salicaceae; Populus.
1 (bases 1 to 592)

REFERENCE
AUTHORS
RAI, S., COOPER, D., KOLSOVA, N., ODDY, C., BUTTERFIELD, Y.,
KIRKPATRICK, R., LIU, J., PALMQUIST, D., STOUT, J., BARBER, S., YANG, G.,
BABAKOFF, R., BROWN, J., OLSEN, T., KONG, D., RITLAND, C.E., SIDDIQUI, A.,
MAYO, M., MORAN, J., OLSEN, T., KONG, D., RITLAND, C.E., SIDDIQUI, A.,
HOLT, R., JONES, S., MARRA, M., ELLIS, B.E., DOUGLAS, C., RITLAND, K., and
Bohlmann, J.

TITLE
The poplar transcriptome: Analysis of expressed sequence tags from
multiple cDNA libraries
JOURNAL
COMMENT
Contact: Joerg Bohlmann
Genome BC forest genomics program
University of British Columbia
UBC Biotechnology Laboratory, 6174 University Boulevard, Rm. 237,
Vancouver, British Columbia, Canada, V6T 1Z3
Tel: 1-604-822-0282
Fax: 1-604-822-6097
Email: bohlmann@interchange.ubc.ca
Plate: WS0118 row: F column: 22
High quality sequence stop: 592
POLYA=yes

FEATURES

source
location/Qualifiers
1..592
/organism="Populus balsamifera subsp. trichocarpa"
/mol_type="mRNA"
/catalytic="383-2499 (Nisqually-1)"
/sub_species="trichocarpa"
/db_xref="taxon:3694"
/clone="WS0118_P22"
/sex="female"
/lab_host="E. coli DH10B T1 phage resistant cells"
/clone_id="PF-P-FL-A-2"
/note="Vector: pluscript II SK (+) XR; Site 1: SacI (5'
end of cDNA); Site 2: XhoI (3' end of cDNA); Ploem and

cambium from 8 year old trees harvested within the Boise Cascade region of Washington state on May 15th, 2001. cDNA was prepared from 20 micrograms of mRNA according to the full-length cDNA library construction method described by Carninci P. et al. (2000). Genome Research 10(10):1617-1630 and directionally ligated into the pBluescript II SK (+) XR vector digested with SstI (5' end) and XhoI (3'). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation."

ORIGIN

Query Match 92.0%; Score 18.4; DB 7; Length 592;
Best Local Similarity 95.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGAAGTGGGCAATCGT 20
|||||
545 TCGAAGTGGGCAATCGT 526

RESULT 2

AZ564202/c 633 bp DNA linear GSS 20-NOV-2000
LOCUS RPCI-23-208N19.TV RPCI-23 Mus musculus genomic clone
DEFINITION RPCI-23-208N19, genomic survey sequence.
ACCESSION AZ564202
VERSION AZ564202.1 GI:11244022
KEYWORDS GSS.

SOURCE

Mus musculus (house mouse)
Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
1 (bases 1 to 633)

REFERENCE

Zhao, S., Nierman, W., Feldbylum, T., Malek, J., Shateman, S., Akhtar, B., Levine, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Frazer, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)

AUTHORS

Other GSS: RPCI-23-208N19.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

TITLE

Mouse BAC End Sequences from Library RPCI-23

JOURNAL

Unpublished (1999)

COMMENT

Email: szhae@igf.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pleher de Jong (pleher@edj.med.bufo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.bufo.edu/orderingframe.htm>) or from Reseach Genetics (<http://resgen.com>). BAC end page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
Plate: 208 row: N column: 19
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers

1..633
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-208N19"
/sex="Female"
/lab_host="DH10B"
/clone_11b="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1: SCORI; Site_2: ECORI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of ECORI and SCORI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the SCORI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN

Query Match 87.0%; Score 17.4; DB 8; Length 633;
Best Local Similarity 94.7%; Pred. No. 4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGAAGTGGGCAATCGT 19
|||||
491 TCGAAGTGGGCAATCGT 473

RESULT 3

CK124820 559 bp mRNA linear EST 01-MAR-2004
LOCUS BS1824110H21 BS1824 Hordeum vulgare subsp. vulgare cDNA clone
DEFINITION MPMP2010H2110 5-PRIME, mRNA sequence.

ACCESSION

CK124820
CK124820.1 GI:44807822

VERSION

EST.

KEYWORDS

Hordeum vulgare subsp. vulgare

SOURCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Hordeum.

ORGANISM

Hordeum vulgare subsp. vulgare

REFERENCE

Kramer, A., Fellner, T., Possling, A., Radchuk, V., Weschke, W., Buerkle, L. and Kersten, B.

AUTHORS

Application of the protein microarray technology for the identification of expression library derived target proteins for barley protein kinase CK2

TITLE

Unpublished (2003)

JOURNAL

Contact: Birgit Kersten* and Winfriede Weschke**
*Plant Protein Chip Group, Department Lebrach, **Department Molecular Genetics, Gene Expression Group
**Max-Planck-Institute for Molecular Genetics, **Institute of Plant Genetics and Crop Plant Research Gatersleben
*Institut. 73, D-14195 Berlin, Germany, **Corrensstrasse 3, D-06466 Gatersleben, Germany
Tel.: **49(0)30/84131648, **49(0)394825500
Fax: **49(0)30/84131128, **49(0)394825237
Email: *kersten@molgen.mpg.de, **weschke@ipk-gatersleben.de
Insert Length: 559 Std Error: 0.00
Plate: 10 row: H column: 21
Seq primer: POB65.

FEATURES

Location/Qualifiers

1..559
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="barke"
/sub_species="vulgare"
/db_xref="GABI:945905"
/db_xref="taxon:112509"
/clone="MPMP2010H2110"
/tissue_type="embryo sac"
/dev_stage="0-10 DAF (days after flowering)"
/lab_host="E. coli, SCS-1/pSE111"
/clone_11b="BS1824"
/note="Vector: pOR30NST (AP074376); Site_1: SalI; Site_2: NotI; 0-10 DAF (days after flowering). cDNA synthesis using pBluescript II XR cDNA-library construction kit (Stratagene) with an oligo(dT)-primer containing NotI restriction site and a SalI adapter (Invitrogen). The main library of 21500 clones was rearranged into the sublibrary BS1824 containing 4100 putative expression clones. Note: Due to a cloning artefact caused by the kit, in most cases the SalI site is NOT present, as well as the SalI Adapter used for cloning. To excise the insert, restriction sites upstream SalI should be used (e.g. BamHI). Average insert size is 1 kb. Library generation and sequencing was granted in context of GABI; data are also accessible at <https://gabi.rzp.de>"

ORIGIN

Query Match 85.0%; Score 17; DB 7; Length 559;

Best Local Similarity 100.0%; Pred. No. 6.3e+02; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GAAGGTGGGCAAAATCGT 19
|||||
540 GAAGGTGGGCAAAATCGT 556

RESULT 4

BF627004/c

LOCUS 790 bp mRNA linear EST 22-OCT-2001

DEFINITION HVSMED001P03f Hordeum vulgare seedling shoot EST library
HVCNA0002 (Dehydration stress) Hordeum vulgare subsp. vulgare cDNA
clone HVSMED001P03f, mRNA sequence.

ACCESSION BF627004

VERSION BF627004.2 GI:13088570

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL
COMMENT

Unpublished (2001)
On Dec 19, 2000 this sequence version replaced gi:11891162.

Contact: Wang RA

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rtwing@clemson.edu

Total hg bases = 433

Seq primer: AATTAACCTCACTAAGGG

High quality sequence start: 4

High quality sequence stop: 486.

FEATURES

source

1..790

Location/Qualifiers

/organism="Hordeum vulgare subsp. vulgare"

/mol_type="mRNA"

/cultivar="Morex"

/sub_species="vulgare"

/db_xref="taxon:112509"

/clone="HVSMED001P03f"

/tissue_type="Seedling shoot"

/lab_host="TJUC121"

/clone_lib="Hordeum vulgare seedling shoot EST library

HVCNA0002 (Dehydration stress)"

/note="vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;

Seeds were surface sterilized then germinated under axenic

conditions in the dark at room temperature in covered

with water, nystatin and cefotaxime in covered

crystallization dishes. Five-day old seedlings were

incubated at 90° RH for 24 hr. Shoots were then harvested,

total RNA was prepared, poly(A) RNA was purified, one

primary unamplified cDNA library was made, 600000 plv were

in vivo excised to give Bluescript SK(-) cDNA phagemids.

These steps were performed in the TV Close laboratory at

the University of California, Riverside (Choi, Close,

Fenton). Phagemids were plated and picked at the Clemson

University Genomics Institute (CUGI) (Begum, Palmer,

Friesch, Atkins and Wing). Plasmid DNA preparations, DNA

sequencing and sequence analysis were performed at CUGI

(Wing, Yu, Friesch, Henry, Simmons, Oates, Rambo, Main).

The sequence has been trimmed to remove vector sequence

and contains a minimum of 100 bases of phred value 20 or

ORIGIN

Query Match 85.0%; Score 17; DB 2; Length 790;

Best Local Similarity 100.0%; Pred. No. 6.7e+02; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GAAGGTGGGCAAAATCGT 19
|||||
519 GAAGGTGGGCAAAATCGT 503

RESULT 5

CG049330/c 391 bp DNA linear GSS 19-AUG-2003

LOCUS PU1FQ44TD ZM 0.6.1.0_KB Zea mays genomic clone ZMMBRC04572H15,

DEFINITION genomic survey sequence.

ACCESSION CG049330

VERSION CG049330.1 GI:33921510

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2003)

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@cigr.org

Seq primer: TP

Class: sheared ends.

Location/Qualifiers

1..391

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMMBRC04572H15"

/clone_lib="ZM 0.6.1.0_KB"

/note="vector: pCR4-TOPO, Site 1: EcoRI; 0.6-1.0 kb high

cot selected genomic DNA library"

ORIGIN

Query Match

Best Local Similarity 90.0%; Pred. No. 7.6e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGAAGTGGGCAAAATCGTG 20

|||||

306 TTGAAGTGGGCAAAATCGTG 287

|||||

CG049330

LOCUS CL219748/c

DEFINITION

ZMMBRC0447H13f ZMMBRC Zea mays genomic clone ZMMBRC0447H13 5',

genomic survey sequence.

RESULT 6

CL219748/c

452 bp DNA linear GSS 08-JAN-2004

ZMMBRC0447H13f ZMMBRC Zea mays genomic clone ZMMBRC0447H13 5',

genomic survey sequence.

above. For more details on library preparation and

sequence analysis see

http://www.genome.clemson.edu/projects/barley. To order

this clone see http://www.genome.clemson.edu/orders Also

see Close TJ, Wing R, Kleinof A, Wise R (2001)

Genetically and physically anchored EST resources for

barley genomics. Barley Genetics Newsletter 31:29-30.

(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ACCESSION CL219748
 VERSION CL219748.1 GI:40752422
 KEYWORDS GSS.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 452)
 Bharti,A.K., Young,S., Kawchok,S., Keizer,G., Bronzino,A.C., Zouavezy,V., Fuks,G., Yu,Y., Wang,R. and Messing,J.
 Sequencing of the maize genome at PGIR (2003c)
 Unpublished (2003)
 CONTACT: Bharti,A.K.
 Dr. Joachim Messing's lab
 The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers University
 190 Frelinghuysen Road, Piscataway, NJ 08854, USA
 Tel: 732 445 3801
 Fax: 732 445 5735
 Email: bharti@waksman.rutgers.edu
 Seq primer: 77
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1..452
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone="ZMMBc047H13"
 /lab_host="E. coli DH10B"
 /clone_1lb="ZMMBc"
 /note="Vector: pTRABAC1.3; Site_1: BamHI; Site_2: BamHI"

ORIGIN
 Query Match 84.0%; Score 16.8; DB 9; Length 452;
 Best Local Similarity 90.0%; Pred. No. 7.8e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCGTG 20
 |||||
 131 TCGAAGTGGTCAATCGTG 112

RESULT 7
 CC803612 511 bp DNA linear GSS 14-JUL-2003
 DEFINITION ih17b02, WGS-Zmaysu (DH10B or JM107MA2) Zea mays genomic clone
 CC803612
 VERSION CC803612
 KEYWORDS GSS.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 511)
 Rabinowitz,P.D., O'Shaughnessy,A.L., Palmer,L., Balija,V., Dedolia,N., Katzenburger,F., King,J., Miller,B., Muller,S., Nacimento,L., Zultavern,T., McCombie,W.R. and Martienssen,R.A.
 Genomic shotgun sequences from Zea mays (unfiltered)
 Unpublished (2003)
 CONTACT: W. Richard McCombie
 Lita Annenberg Hazen Genome Sequencing Center
 Cold Spring Harbor Laboratory
 PO Box 100, Cold Spring Harbor, NY 11724, USA
 Tel: 516 367 8884
 Fax: 516 367 8874
 Email: mcombie@cshl.org
 Plate: ih17 row: b column: 02
 Seq primer: -21M3UnivRev
 Class: Shotgun

TITLE
 JOURNAL
 COMMENT

High quality sequence stop: 511.
 Location/Qualifiers
 1..511
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone="ih17b02"
 /lab_host="DH10B or JM107MA2"
 /clone_1lb="WGS-Zmaysu (DH10B or JM107MA2)"
 /note="Organ: immature ears; Site_1: Xba I; Site_2: Xba I;
 The vector was digested with XbaI and one nucleotide was added by fill in in the recessive 3' end. The genomic DNA was nebulized, end repaired, adaptor ligated and size fractionated using sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (.x/y reads in M13mp19, .b/g reads in pUC19). The ligation was transformed into DH10B or JM107MA2."

ORIGIN
 Query Match 84.0%; Score 16.8; DB 9; Length 511;
 Best Local Similarity 90.0%; Pred. No. 8e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCGTG 20
 |||||
 278 TCGAAGTGGTCAATCGTG 297

RESULT 8
 BU098971 548 bp mRNA linear EST 01-OCT-2003
 LOCUS BU098971 NIBB Mochii normalized Xenopus early gastrula library
 DEFINITION Xenopus laevis CDNA clone XL148K21 3', mRNA sequence.
 BU098971
 VERSION BU098971
 KEYWORDS EST.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae; Xenopus; Xenopus.
 1 (bases 1 to 548)
 Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and Kohara,Y.
 Expressed genes in X. laevis embryo
 Unpublished (2001)
 CONTACT: Tadasu Shin-I
 Center for Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@gens.nig.ac.jp
 The information of this clone is available through the following URL.
 http://xenopus.nibb.ac.jp.
 Location/Qualifiers
 1..548
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="XL148K21"
 /tissue_type="whole embryo"
 /dev_stage="stage 10.5"
 /clone_1lb="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN
 Query Match 84.0%; Score 16.8; DB 4; Length 548;
 Best Local Similarity 90.0%; Pred. No. 8e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TCGAAGTGGGCAATCGTG 20
 DB 474 TCGAAGTGGGCAATCGG 493

RESULT 9
 B0097494/c 553 bp mRNA linear EST 20-DEC-2002
 LOCUS ph03d06.y2 Oostertagia oostertagi L3 PAMP1 v1 Oostertagia oostertagi
 DEFINITION cDNA 5' similar to TR:Q22371 Q22371 T10B10.4 PROTEIN. [1] ; mRNA
 sequence.

ACCESSION B0097494
 VERSION B0097494
 KEYWORDS EST
 SOURCE B0097494.1 GI:20130478

ORGANISM Oostertagia oostertagi
 Oostertagia oostertagi
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
 Trichostrongyloidea; Haemonchidae; Oostertagiinae; Oostertagia.
 REFERENCE 1 (bases 1 to 553)
 McCarter,J., Clifton,S., Chiappelli,B., Pape,D., Martin,J.,
 Wyllie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
 Bowers,Y., Gibbons,M., Ritzer,B., Bennett,J., Franklin,C.,
 Tsagaris,I., R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
 Underwood,K., Stepcoe,M., Allen,M., Person,B., Swaller,T.,
 Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
 McCann,R., Waterston,R. and Wilson,R.
 The Washington Univ. Nematode EST Project, 1999
 Unpublished (1999)

TITLE JOURNAL
 COMMENT Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu

The library was constructed by Claire Murphy and Dr. James McCarter
 at Washington University, St. Louis. DNA Sequencing by: Washington
 University Genome Sequencing Center
 High quality sequence stop: 415.
 Location/Qualifiers

FEATURES
 source 1..553
 /organism="Oostertagia oostertagi"
 /mol_type="mRNA"
 /db_xref="taxon:6317"
 /dev_stage="L3"
 /lab_host="DH10B"
 /clone_lib="Oostertagia oostertagi L3 PAMP1 v1"
 /note="Vector: PAMP1 (Gibco); Site 1: NotI; Site 2: SalI;
 ***WARNING: Subsequent examination of these samples has
 revealed the presence of an additional Trichostrongyloidea
 cattle nematode, Cooperia oncophora. Sequences in this
 library may derive from either Oostertagia or Cooperia.***
 The library was constructed by Claire Murphy and Dr. James
 McCarter at Washington University, St. Louis. The cDNA was
 made by using Dynabead oligo-dT priming (DynaI). PCR based
 library using a modified protocol from the SMART PCR cDNA
 Synthesis Kit from Clontech. Directionally cloned into the
 UDG sites of PAMP1. Nematodes were provided by Dr. Louis
 Gasbarré of the USDA, Beltsville, MD
 (lgasbarr@nri.barc.usda.gov). Third stage exsheathed
 larvae were collected from 14 day fecal-sphagnum moss
 cultures of Oostertagia eggs. The larvae were recovered by
 overnight passage on a Baermann apparatus, and then
 cleaned by passage through a 20 micron nylon mesh. The
 larvae were then subjected to a treatment with 1.25%
 chlorox to induce excystation. The larvae were washed with
 5 changes of PBS and then pelleted and snap frozen in
 liquid nitrogen."

ORIGIN

Query Match 84.0%; Score 16.8; DB 5; Length 553;
 Best Local Similarity 90.0%; Pred. No. 8.1e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TCGAAGTGGGCAATCGTG 20
 DB 525 TCGAAGTGGGCAATCGTG 506

RESULT 10
 A2471085/c 607 bp DNA linear GSS 04-OCT-2000
 LOCUS 1M0285D02R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 DEFINITION clone UUGC1M0285D02 R, genomic survey sequence.

ACCESSION A2471085
 VERSION A2471085
 KEYWORDS GSS
 SOURCE A2471085.1 GI:10629210

ORGANISM Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 607)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
 Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
 Niederhausern,A. and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)

TITLE JOURNAL
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: dunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00
 Plate: 0285 row: D column: 02
 Seq primer: CACACAGGAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 607.
 Location/Qualifiers

FEATURES
 source 1..607
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0285D02"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /note="Vector: PMD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD2 (91473214[gb]AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

ORIGIN

Query Match 84.0%; Score 16.8; DB 8; Length 607;
 Best Local Similarity 90.0%; Pred. No. 8.2e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCTGTG 20
|||||
Db 58 TCGACGTGGGCAAGCGTGTG 39

RESULT 11
LOCUS BJ460007
DEFINITION BJ460007 K. Sato unpublished cDNA library, cv. Akashinriki
vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
baak39b17 3', mRNA sequence.
ACCESSION BJ460007
VERSION BJ460007.1 GI:21138532
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 608)
Sato, K., Saisho, D. and Takeda, K.
Barley EST sequencing project in NIG and Okayama Univ
JOURNAL Unpublished (2002)
CONTACT: Tadasu Shin-1
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..608
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Akashinriki"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="baak39b17"
/tissue_type="leaves"
/dev_stage="vegetative stage"
/clone_lib="K. Sato unpublished cDNA library, cv.
Akashinriki vegetative stage leaves"

ORIGIN
Query Match 84.0%; Score 16.8; DB 4; Length 608;
Best Local Similarity 90.0%; Pred. No. 8.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCTGTG 20
|||||
Db 142 TCGAAGCGCGCAATCTTGTG 161

RESULT 12
LOCUS AV910592/c 627 bp mRNA linear EST 18-JAN-2002
DEFINITION AV910592 K. Sato unpublished cDNA library, cv. Akashinriki
vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
baak1a11 5', mRNA sequence.
ACCESSION AV910592
VERSION AV910592.1 GI:18206357
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 627)
Sato, K., Saisho, D. and Takeda, K.
Barley EST sequencing project in NIG and Okayama Univ
JOURNAL Unpublished (2002)
CONTACT: Tadasu Shin-1

ORIGIN
Query Match 84.0%; Score 16.8; DB 4; Length 608;
Best Local Similarity 90.0%; Pred. No. 8.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..627
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Akashinriki"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="baak1a11"
/tissue_type="leaves"
/dev_stage="vegetative stage"
/clone_lib="K. Sato unpublished cDNA library, cv.
Akashinriki vegetative stage leaves"

ORIGIN
Query Match 84.0%; Score 16.8; DB 1; Length 627;
Best Local Similarity 90.0%; Pred. No. 8.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCTGTG 20
|||||
Db 600 TCGAAGCGCGCAATCTTGTG 581

RESULT 13
LOCUS BJ462359 642 bp mRNA linear EST 23-MAY-2002
DEFINITION BJ462359 K. Sato unpublished cDNA library, cv. Akashinriki
vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
baak24n23 3', mRNA sequence.
ACCESSION BJ462359
VERSION BJ462359.1 GI:21140866
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 642)
Sato, K., Saisho, D. and Takeda, K.
Barley EST sequencing project in NIG and Okayama Univ
JOURNAL Unpublished (2002)
CONTACT: Tadasu Shin-1
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..642
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Akashinriki"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="baak24n23"
/tissue_type="leaves"
/dev_stage="vegetative stage"
/clone_lib="K. Sato unpublished cDNA library, cv.
Akashinriki vegetative stage leaves"

ORIGIN
Query Match 84.0%; Score 16.8; DB 4; Length 642;
Best Local Similarity 90.0%; Pred. No. 8.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCGAAGTGGGCAATCTGTG 20

Db 13 TCGAAGCGCGCAATCTTG 32

RESULT 14
BJ452459/c

DEFINITION BJ452459 K. Sato unpublished cDNA library, cv. Akashiniriki
vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
baak39b17 5', mRNA sequence.

ACCESSION BJ452459

VERSION BJ452459.1 GI:21131047

KEYWORDS EST.

SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 646)
Sato, K., Saiho, D. and Takeda, K.
Barley EST sequencing project in NIG and Okayama Univ
Unpublished (2002)
Contact: Tadau Shin-i
Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source
1..646
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Akashiniriki"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="baak39b17"
/tissue_type="leaves"
/dev_stage="vegetative stage"
/clone_lib="K. Sato unpublished cDNA library, cv.
Akashiniriki vegetative stage leaves"

ORIGIN

Query Match 84.0%; Score 16.8; DB 4; Length 646;
Best Local Similarity 90.0%; Pred. No. 8.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGAAGTGGCGCAATCTTG 20
|||||
Db 482 TCGAAGCGCGCAATCTTG 463

RESULT 15
AV920519 658 bp mRNA linear EST 18-JAN-2002
LOCUS AV920519
DEFINITION AV920519 K. Sato unpublished cDNA library, cv. Haruna Nijo
germination shoots Hordeum vulgare subsp. vulgare cDNA clone
bags14g03 3', mRNA sequence.

ACCESSION AV920519

VERSION AV920519.1 GI:18216298

KEYWORDS EST.

SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 658)
Sato, K., Saiho, D. and Takeda, K.
Barley EST sequencing project in NIG and Okayama Univ
Unpublished (2002)
Contact: Tadau Shin-i
Center For Genetic Resource Information
National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1..658
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Haruna Nijo"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="bags14g03"
/tissue_type="shoots"
/dev_stage="germination"
/clone_lib="K. Sato unpublished cDNA library, cv. Haruna
Nijo germination shoots"

ORIGIN

Query Match 84.0%; Score 16.8; DB 1; Length 658;
Best Local Similarity 90.0%; Pred. No. 8.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGAAGTGGCGCAATCTTG 20
|||||
Db 34 TCGAAGCGCGCAATCTTG 53

Search completed: November 7, 2005, 09:18:42
Job time : 2008.85 secs

This Page From Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:11:16 ; Search time 746.66 seconds
(without alignments)
1233.024 Million cell updates/sec

Title: US-10-777-131A-4

Perfect score: 19

Sequence: 1 tggggcttcgctggaagtg 19

Scoring table: IDENTITY_NUC

Searched: 4708233 seqs, 2422767955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ha: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_gts: *
12: gb_gy: *
13: gb_un: *
14: gb_vi: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	100.0	19	6	AX770106 Sequence
2	19	100.0	1144	6	AX770122 Sequence
3	19	100.0	2406	6	AX770134 Sequence
4	19	100.0	2409	6	BD273957 Sequences
5	19	100.0	2409	6	AX005499 Sequences
6	19	100.0	2451	6	BD273958 Sequences
7	19	100.0	2451	6	AX005501 Sequences
8	19	100.0	2806	6	AX770124 Sequence
9	19	100.0	35336	1	MSGY367
10	19	100.0	103837	1	AE000516_43
11	19	100.0	244800	1	BX842584 Mycobacte
12	19	100.0	278492	1	BX248347 Mycobacte
13	19	100.0	278492	1	CR847886 Mus muscu
14	19	100.0	278492	1	AC115304 Mus muscu
15	19	100.0	278492	1	AC115304 Mus muscu
16	19	100.0	278492	1	AC115304 Mus muscu
17	19	100.0	278492	1	AC115304 Mus muscu
18	19	100.0	278492	1	AC115304 Mus muscu
19	19	100.0	278492	1	AC115304 Mus muscu

20	17	89.5	235148	2	AC113808
21	17	89.5	259243	1	AC106670 Rattus no
22	17	89.5	347636	1	BX842653 Bdellovib
23	16.4	86.3	2044	10	BC020122
24	16.4	86.3	110000	2	AC095736_1
25	16.4	86.3	110000	2	AC127735_1
26	16.4	86.3	117328	6	C0870156 Sequence
27	16.4	86.3	121813	2	AC147587
28	16.4	86.3	139976	2	AC145482
29	16.4	86.3	167739	2	AC020547
30	16.4	86.3	180000	2	AC102737
31	16.4	86.3	191887	2	CR407563
32	16.4	86.3	236186	2	AC094066
33	16.4	86.3	262582	2	AC099145
34	16.4	86.3	270991	2	AC136210
35	16.4	86.3	279311	2	AC131439
36	16.4	86.3	334051	2	AC123374
37	16	84.2	116951	9	AC114788
38	16	84.2	159377	2	AC147522
39	16	84.2	168277	2	AC147934
40	16	84.2	174024	9	AL583827
41	16	84.2	185316	2	AC134386
42	16	84.2	229063	2	AC147751
43	16	84.2	300162	2	AC149441
44	16	84.2	310613	1	AE016875
45	15.8	83.2	201	11	BV179021

ALIGNMENTS

RESULT 1	AX770106	Sequence 4 from Patent WO03016562.	19 bp	DNA	linear	PAT 02-JUL-2003
LOCUS	AX770106					
DEFINITION	AX770106					
ACCESSION	AX770106					
VERSION	AX770106.1	GI:32437684				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						
ORIGIN						
Query Match						
Best Local Similarity						
Matches						
19; Conservative						
0; Mismatches						
0; Indels						
0; Gaps						
0;						
RESULT 2	AX770122	Sequence 20 from Patent WO03016562.	1144 bp	DNA	linear	PAT 02-JUL-2003
LOCUS	AX770122					
DEFINITION	AX770122					
ACCESSION	AX770122					
VERSION	AX770122.1	GI:32437700				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						
ORIGIN						
Query Match						
Best Local Similarity						
Matches						
19; Conservative						
0; Mismatches						
0; Indels						
0; Gaps						
0;						

REFERENCE 1 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

AUTHORS Gicquel,B.

TITLE Compositions and methods for detecting multidrug resistant strains of M. tuberculosis having mutations in genes of the mult family

JOURNAL Patent: WO 03016562-A 20 27-FEB-2003;

INSTITUT PASTEUR (FR)

FEATURES Location/Qualifiers

source 1..1144

/organism="Mycobacterium tuberculosis"

/mol_type="unassigned DNA"

/db_xref="taxon:1773"

ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 1144;

Best Local Similarity 100.0%; Pred. No. 37;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19

Db 1143 TGGGGTTCGCTGGAAGTGG 1125

RESULT 3

AX770134/c 2406 bp DNA linear PAT 02-JUL-2003

LOCUS Sequence 32 from Patent WO03016562.

DEFINITION AX770134

ACCESSION AX770134.1 GI:32437712

VERSION

KEYWORDS

SOURCE

ORGANISM

Myobacterium tuberculosis

Mycobacterium tuberculosis

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium

tuberculosis complex.

REFERENCE 1

AUTHORS Gicquel,B.

TITLE Compositions and methods for detecting multidrug resistant strains of M. tuberculosis having mutations in genes of the mult family

JOURNAL Patent: WO 03016562-A 32 27-FEB-2003;

INSTITUT PASTEUR (FR)

FEATURES Location/Qualifiers

source 1..2406

/organism="Mycobacterium tuberculosis"

/mol_type="unassigned DNA"

/db_xref="taxon:1773"

ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 2406;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19

Db 200 TGGGGTTCGCTGGAAGTGG 182

RESULT 4

BD273957/c 2409 bp DNA linear PAT 17-JUL-2003

LOCUS Sequence nucleic acid from polypeptides exportes from

DEFINITION mycobacteries, vector the complement and applications at diagnostic

and the prevention from the tuberculose.

BD273957

BD273957.1 GI:33083725

VERSION

KEYWORDS

SOURCE

ORGANISM

Myobacterium tuberculosis

Mycobacterium tuberculosis

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium

tuberculosis complex.

REFERENCE 1 (bases 1 to 2409)

AUTHORS Gicquel,B., Portnoi,D., Lim,E., Pellicic,V., Guigueno,A. and

Salmoniere,Y.G.D.L.

TITLE Sequences nucleic acid from polypeptides exportes from

mycobacteries, vector the complement and applications at diagnostic

and the prevention from the tuberculose

Patent: JP 2002534956-A 181 22-OCT-2002;

JOURNAL INSTITUT PASTEUR

COMMENT OS Mycobacterium tuberculosis

PN JP 2002534956-A/181

PD 22-OCT-2002

PF 14-AUG-1998 JP 2000509849

PR 14-AUG-1997 FR 97/10404,11-SEP-1997 FR 97/11325 PI

BRIGITTE GICQUEL,DENIS PORTNOI,ENG-MONG LIM,VLADIMIR PELICIC, PI

AGNES GUIGUENO,YVES GOGUET DE LA SALMONIERE

PC C12N15/09,A61K39/04,A61K48/00,A61P11/00,A61P11/04,C07K14/35,

PC C07K16/12,

PC C07K19/00,C12N1/21,C12P21/02,C12Q1/02,C12Q1/68,G01N33/50, PC

G01N33/53//

PC (C12N15/09,C12R1:32),C12N15/00,(C12N15/00,C12R1:32) CC SEQ

ID No. 36D

FI Key

FT CDS Location/Qualifiers

(1)..(2406).

FEATURES Location/Qualifiers

source 1..2409

/organism="Mycobacterium tuberculosis"

/mol_type="genomic DNA"

/db_xref="taxon:1773"

ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 2409;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19

Db 200 TGGGGTTCGCTGGAAGTGG 182

RESULT 5

AX005499/c 2409 bp DNA linear PAT 24-AUG-2000

LOCUS Sequence 618 from Patent WO9909186.

DEFINITION AX005499

ACCESSION AX005499.1 GI:9928645

VERSION

KEYWORDS

SOURCE

ORGANISM

Myobacterium tuberculosis

Mycobacterium tuberculosis

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium

tuberculosis complex.

REFERENCE 1

AUTHORS Portnoi,D. and Guigueno,A.

TITLE Polypeptide nucleic sequences exported from mycobacteria, vectors

comprising same and uses for diagnosing and preventing tuberculosis

Patent: WO 9909186-A 618 25-FEB-1999;

JOURNAL PORTNOI DENIS (FR); GUIGUENO AGNES (FR)

FEATURES Location/Qualifiers

source 1..2409

/organism="Mycobacterium tuberculosis"

/mol_type="unassigned DNA"

/db_xref="taxon:1773"

1..2409

/note="unnamed protein product; SEQ ID NO 36D"

/codon_start=1

/transl_table=1

/protein_id="CAC05103.1"

/db_xref="GI:9928646"

/translation="MTALQLGMAALARYTSAGVAGIGMALTPVSAAPHALAGRSP

TPFVQVRIDQVTPVITTSSEPHVTSGVTNTGDRPVQDVVRLHHAAYTSALR

TSIAGGDMQGPADFLTVAPELDRGAGFTLSAPSLAPSAVNOGIVPVLVN

VNGRPDYGPAPRLDARFLRVVGVPPQADIDFSAVAPEPTAPVITMIMPLARPR

LPAPAGTIVPRLVDDLLANSLANGRLDILLSAABFATNREVDPDGAVGRALCLAI

DPDLITVNTAMTGIVSDSPDGAALPPTPTHTPGTGOAASWILRLRTVHTCVT
PLPFAQADLDLORVNDRLSALITISPADIVDRILVDSSTRGATVLPDGPITGRAIN
ILSTHGNVVAVAADSPREBOGSSOISALIPAPRSDSRVVAAPDPRAVGAALA
AAGNPTPTTLPDPSLFRIRAHESITARODALGAMLRSLBNAPAPPTOLVPPASW
SLASDDAQVILITALTATIRSGLAVERPAPVADAAATEPEPEAPSAAGRND
ITTOIGGVAVILMTLSALTITDTRGLGVAVTADMRALSCQSAPDPRNLAC
ORLAVGKIDLDLFGAVTIVNGGSYTTATSHSPPLAHNGLAVERIVRLQVDPAPG
MTVADVGOIBLPGVPLPRVPIEVNFTQRVAVDSLRTPDGVALGEPVRLSHSNAYG
KVLPAITLSAAAVLVTLAAGRILMHRFRQGPBRADRDRLPTGKHAPORRAVASNDDE
KHRV"

ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 2409;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGGGTTCGCTGGAAGTGG 19
DB 200 TGGGGTTCGCTGGAAGTGG 182

RESULT 6

BD273958/c

BD273958 2451 bp DNA linear PAT 17-JUL-2003
LOCUS

DEFINITION Sequences nucleic acid from polypeptides exportes from
mycobacteries, vector the complement and applications at diagnostic
and the prevention from the tuberculose.

ACCESSION BD273958
VERSION BD273958.1 GI:33083726

KEYWORDS

JP 2002534956-A/182.

SOURCE

Mycobacterium tuberculosis

ORGANISM

Mycobacterium tuberculosis

REFERENCE

AUTHORS

TITLE

Sequences nucleic acid from polypeptides exportes from
mycobacteries, vector the complement and applications at diagnostic
and the prevention from the tuberculose

JOURNAL

INSTITUT PASTEUR

COMMENT

OS Mycobacterium tuberculosis
PN JP 2002534956-A/182

PD 22-OCT-2002

PF 14-AUG-1998 JP 2000509849

PR 14-AUG-1997 FR 97/10404,11-SEP-1997 FR 97/11325 PI

BRIGITTE GICQUEL, DENIS PORTNOI, ENG-MONG LIM, VLADIMIR PELICIC, PI

AGNES GUIGUENO, YVES GOGUET DE LA SALMONIERE

PC C12N15/09, A61K39/04, A61K48/00, A61P11/00, A61P31/04, C07K14/35,

PC C07K16/12, PC C07K19/00, C12N1/21, C12P21/02, C12Q1/02, C12Q1/68, G01N33/50, PC

G01N33/53//

PC (C12N15/09, C12R1/32), C12N15/00, (C12N15/00, C12R1/32) CC SEQ

ID No. 36F

FH Key

FT CDS

Location/Qualifiers (4) (2448).

Location/Qualifiers

1. 2451

/organism="Mycobacterium tuberculosis"

/mol_type="genomic DNA"

/db_xref="taxon:1773"

ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 2451;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGGGTTCGCTGGAAGTGG 19
DB 242 TGGGGTTCGCTGGAAGTGG 224

RESULT 7

AX005501/c

LOCUS

AX005501 2451 bp DNA linear PAT 24-AUG-2000

DEFINITION

Sequence 620 from Patent WO9909186.

AX005501

ACCESSION

AX005501.1 GI:9928647

VERSION

KEYWORDS

SOURCE

ORGANISM

Mycobacterium tuberculosis

Bacteria; Actinobacteria; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium

tuberculosis complex.

REFERENCE

AUTHORS

TITLE

JOURNAL

CDS

Location/Qualifiers

1. 2451

/organism="Mycobacterium tuberculosis"

/mol_type="unassigned DNA"

/db_xref="taxon:1773"

/note="unamed protein product; SEQ ID NO 36F"

/codon_start=1

/transl_table=1

/protein_id="CAC05104.1"

/db_xref="GI:9928648"

/translation="LSTGSAORSRAGAVTALQLGMAALARTSAIGVAGGMAITVP

SAAPHALGSPSPFPVORIDQVTPDVTSSPHVTVSGTVMNTGPRPDVWRL

EHAAYVSTPLRTSLDGTQYQPADFLVAREIDRGCAFTLSPLASTRSL

AVNPGIPELVANVNGTDPGAPRLNDRILVVGVPDQATFGSAVAPETTAV

WITMLPLADPRPLAPGPGTAVRVDLNLASLNGRLDILLASAPATNREVD

PDGAVGRALCIAIDPDLITVNTAMTGIVSDSPDGAALPPTPTHTPGTGOAASWT

DLRLTIVHRTCTVPLPFAQADLDLORVNDRLSALITISPADIVDRILVDSSTRGAT

VLPPGGLTGRAINILSTHGNVVAADSPREBOGSSOISALIPAPRSDSRV

AAPDPRAVGAALAGTPTPTTLPDPSLFRIRAHESITARODALGAMLRSLBNAP

APRTQIIVPPASMSLASDDAQVILITALTATIRSGLAVERPAPVADAAATEPEPE

GAYSAAGRENDITTOIGGVAVILMTLSALTITDTRGLGVAVTADMRALSCQSAP

DPRNLACORLAVGKIDLDLFGAVTIVNGGSYTTATSHSPPLAHNGLAVERIVRL

QVDPAPGMTADVGOIBLPGVPLPRVPIEVNFTQRVAVDSLRTPDGVALGEPVRL

SHSNAYGKVLPAITLSAAAVLVTLAAGRILMHRFRQGPBRADRDRLPTGKHA

APORRAVASRDEKHRV"

OY 1 TGGGGTTCGCTGGAAGTGG 19

DB 242 TGGGGTTCGCTGGAAGTGG 224

ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 2451;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGGGTTCGCTGGAAGTGG 19
DB 242 TGGGGTTCGCTGGAAGTGG 224

RESULT 8

AX770124/c

LOCUS

AX770124 2806 bp DNA linear PAT 02-JUL-2003

DEFINITION

Sequence 22 from Patent WO03016562.

AX770124

ACCESSION

AX770124.1 GI:32437702

VERSION

KEYWORDS

SOURCE

ORGANISM

Mycobacterium tuberculosis

Bacteria; Actinobacteria; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium

tuberculosis complex.

REFERENCE

AUTHORS

TITLE

JOURNAL

Compositions and methods for detecting multidrug resistant strains
of M. tuberculosis having mutations in genes of the mutT family
Patent: WO 03016562-A 22 27-FEB-2003;

INSTITUT PASTEUR (FR)
FEATURES
Location/Qualifiers
Source
1..2806
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"

ORIGIN
Query Match 100.0%; Score 19; DB 6; Length 2806;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTGCTGGAAGTGG 19
|||||
400 TGGGGTTGCTGGAAGTGG 382

RESULT 9
MSGY367 35336 bp DNA linear BCT 03-DEC-1996
LOCUS Mycobacterium tuberculosis sequence from clone y367.
DEFINITION
ACCESSION AD000008
VERSION AD000008.1 GI:1702971
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis
Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinomycetales;
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1 (bases 1 to 35336)

REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (11-OCT-1996) L. Du, Genome Therapeutics Corporation, 100
Beaver Street, Waltham, MA, USA, 02154 du@cric.com
GDB:S:1004717.
COMMENT
FEATURES
Source
1..35336
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"
/clone="y367"

ORIGIN
Query Match 100.0%; Score 19; DB 1; Length 35336;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTGCTGGAAGTGG 19
|||||
Db 13574 TGGGGTTGCTGGAAGTGG 13592

RESULT 10
AE000516_43/c
WPCOMMENT
Sequence split into 44 fragments LOCUS AE000516 Accession AE000516
Fragment Name Begin End
AE000516_00 1 110000
AE000516_01 100001 210000
AE000516_02 200001 310000
AE000516_03 300001 410000
AE000516_04 400001 510000
AE000516_05 500001 610000
AE000516_06 600001 710000
AE000516_07 700001 810000
AE000516_08 800001 910000
AE000516_09 900001 1010000
AE000516_10 1000001 1110000
AE000516_11 1100001 1210000
AE000516_12 1200001 1310000
AE000516_13 1300001 1410000
AE000516_14 1400001 1510000
AE000516_15 1500001 1610000

AE000516_16 1600001 1710000
AE000516_17 1700001 1810000
AE000516_18 1800001 1910000
AE000516_19 1900001 2010000
AE000516_20 2000001 2110000
AE000516_21 2100001 2210000
AE000516_22 2200001 2310000
AE000516_23 2300001 2410000
AE000516_24 2400001 2510000
AE000516_25 2500001 2610000
AE000516_26 2600001 2710000
AE000516_27 2700001 2810000
AE000516_28 2800001 2910000
AE000516_29 2900001 3010000
AE000516_30 3000001 3110000
AE000516_31 3100001 3210000
AE000516_32 3200001 3310000
AE000516_33 3300001 3410000
AE000516_34 3400001 3510000
AE000516_35 3500001 3610000
AE000516_36 3600001 3710000
AE000516_37 3700001 3810000
AE000516_38 3800001 3910000
AE000516_39 3900001 4010000
AE000516_40 4000001 4110000
AE000516_41 4100001 4210000
AE000516_42 4200001 4310000
AE000516_43 4300001 4403837

Continuation (44 of 44) of AE000516 from base 4300001 (AE000516 Mycobacterium tuberculosis
Query Match 100.0%; Score 19; DB 1; Length 103837;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTGCTGGAAGTGG 19
|||||
Db 86697 TGGGGTTGCTGGAAGTGG 86679

RESULT 11
BX842584/c
LOCUS
DEFINITION
ACCESSION
BX842584 AL021426 AL022076 AL022120 AL022121 AL123456 Z80343 Z83864
Z94121 Z97188
VERSION
BX842584.1 GI:38490370
KEYWORDS
SOURCE
ORGANISM
Complete genome.
Mycobacterium tuberculosis H37Rv
Mycobacterium tuberculosis H37Rv
Bacteria; Actinobacteria; Actinomycetales;
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCES
AUTHORS
TITLE
JOURNAL
MEDLINE
1 Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jorgensen, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squares, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S., and
Barrell, B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (6685), 537-544 (1998)
JOURNAL
MEDLINE
98295967
9634230
2 Camus, J.C., Pryor, M.J., Medigue, C. and Cole, S.T.
Re-annotation of the genome sequence of Mycobacterium tuberculosis
H37Rv
Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)
22255591

PUBMED REFERENCE	12368430
3 (bases 1 to 244800)	
AUTHORS	Parkhill,J.
TITLE	Direct Submission
JOURNAL	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT	On or before Nov 21, 2003 this sequence version replaced gi:3261511, gi:3256026, gi:3261558, gi:3261559, gi:3261648, gi:3261687, gi:3261736, gi:33261805.
FEATURES	Notes: Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/).
SOURCE	Location/Qualifiers 1. 244800 /organism="Mycobacterium tuberculosis H37Rv" /mol_type="genomic DNA" /strain="H37Rv" /db_xref="taxon:83332" complement(89..1396) /locus_tag="RV3722C" complement(89..1396) /locus_tag="RV3722C" /function="UNKNOWN" /note="RV3722C. (MTV025.070c), len: 435 aa. Conserved hypothetical protein, equivalent to O69513 MLCB2407.14 (alias O9CB73 ML2336, 463 aa) HYPOTHETICAL 46.8 KDA PROTEIN from Mycobacterium leprae (426 aa), FASTA scores: opt: 2505, E(): 8.3e-154, (87.25% identity in 424 aa overlap). Also highly similar to Q9R1J7 DR1579 CONSERVED HYPOTHETICAL PROTEIN from Deinococcus radiodurans (452 aa). FASTA scores: opt: 1652, E(): 3.1e-67, (44.8% identity in 422 aa overlap); and partially similar to Q91371 PA1654 PROBABLE AMINOTRANSFERASE from Pseudomonas aeruginosa (388 aa) FASTA scores: opt: 162, E(): 0.0078, (25.85% identity in 348 aa overlap) and other aminotransferases. TParse score is 0.900. N-terminus extended since first submission (previously 408 aa)." /codon_start=1 /evidence=experimental /transl_table=1 /product="CONSERVED HYPOTHETICAL PROTEIN" /protein_id="CAA18044.2" /db_xref="gi:38490371" /translation="MSFDSLSPOELALAHARHODVYALQGMKTLALDTTRGSPASBOLDLSNQLSLPGDDVDRPBGCTPTVRVGGOHGPGRLPAELIGAVNLTLAGNNSISLELMHDIVAFSLHYGVVDSPPRPIQEQDDIKELCPFGIDRHAFTTMTGIENTPIPTMQLDGEVDVLIIEELVAVDPALIKGMWTVFVGNBSGVYSWEYARLVQMTAADEFYLPMDNAAVAVLTLDLPPRQVVLGLAAAGAPNRRVYASYSKLTTPAGGVSFFGSGIGNLAESYLQYAAKSKSGIDPKVQQLRHLRFGGADGVRLSRKQTLAKRFALVVELVDORLSWYLSKAAWTEPKGGYFSLIDVLPPTARVTLAKVQGLAVTAGASPYRKDPDDKNIRV" /locus_tag="1638 1613. 1638 /gene="trna-ser (GGA) " 1613. 1638 /gene="trna-ser (GGA) " /product="trna-ser" /note="codon recognized: UCC" /anticodon (pos:1647..1649,aa:Ser) 1804. 2568 /locus_tag="RV3723 " 1804. 2568 /locus_tag="RV3723 " 1804. 2568 /function="UNKNOWN" /note="RV3723. (MTV025.071), len: 254 aa. Probable conserved transmembrane protein, with hydrophobic stretches at the N-terminus, and equivalent to O69512 ML2337 MLCB2407.13c PUTATIVE MEMBRANE PROTEIN from Mycobacterium leprae (250 aa), FASTA scores: opt: 1029, E(): 1.2e-44, (64.45% identity in 253 aa overlap). TParse

```

score is 0.900."
/codon_start=1
/tranl_table=1
/product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN"
/protein_id="CA18045.1"
/db_xref="GI:2960147"
/translation="MGRRVAVLTMASFSIGAVLYFEVLPRLWELMGDTGHSIGTGL
RIAGLADGMLPVPFLILRTREDEIGTQDILISMISWIMHVLGALIVGTAISE
VWLSIDAGQMLPGTICGAAAIATVAFPGYLSVVALPPRPDPLKKPKRRRLRR
KKTAKGDPAEBAEBAEBAETELAAQDEBAEVAEPSPSIEPGESEBATEBAALETRA
TAEERGGSLNRRRTGTSHRRRTRGVOVAKVE"
2735. .2977
/gene="cut5a"
/locus_tag="RV3724A"
2735. .2977
/gene="cut5a"
/locus_tag="RV3724A"
/EC_number="3.1.1.-"
/function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE
STRUCTURE OF PLANT CUTICLE)."
/note="RV3724A, (MTV025.072), len: 80 aa. Probable cut5a,
truncated cutinase precursor (EC 3.1.1.-), similar to
N-terminal end of others e.g. O9K87 SERINE ESTERASE
CUTINASE from Mycobacterium avium (220 aa), FASTA scores:
opt: 202, E(): 1.5e-06, (56.45% identity in 62 aa
overlap); Q9XB09|RV2D-RV1758 PROTEIN (FRAGMENT) from
Mycobacterium bovis BCG (143 aa), FASTA scores: opt: 200,
E(): 1.5e-06, (61.4% identity in 57 aa overlap); and
O00298|CUT1_BOTCI|CUT1A CUTINASE PRECURSOR from Botrytis
cinerea (Botryotinia fuckeliana) (202 aa), FASTA scores:
opt: 108, E(): 2.2, (40.4% identity in 52 aa overlap).
Also highly similar to others from Mycobacterium
tuberculosis e.g.
O06181|CUT3_MYCTU|RV3451|MT3557|MTCY3E12.04 PROBABLE
CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 189, E():
1.2e-05, (58.0% identity in 50 aa overlap);
Q50664|CUT2_MYCTU|RV2301|MT2338|MTCY339.08c PROBABLE
CUTINASE PRECURSOR (219 aa), FASTA scores: opt: 172, E():
0.00015, (59.2% identity in 49 aa overlap);
O06793|RV1758|MTCY28.24|Z95890 HYPOTHETICAL 17.9 KDA
PROTEIN (174 aa), FASTA scores: opt: 641, E(): 2.7e-29,
(57.2% identity in 166 aa overlap);
O06191|RV3452|MTY13E12.05; and U00015_11 from Mycobacterium
leprae. BELONGS TO THE CUTINASE FAMILY. Rest of cutinase
ORF continues as RV3724B|CUT5B. frameshifting could occur
near position 4169668. Sequence has been checked but no
errors found."
/codon_start=1
/tranl_table=1
/product="PROBABLE CUTINASE PRECURSOR (FIRST PART) CUT5A"
/protein_id="CA855632.1"
/db_xref="GI:38490372"
/translation="MDVIRMARRLAVAGTAAVTPTEGLSAHPMVMSARPCDEVEV
FARGGEPPGIGSGVGLFVDALRPPGRQVTRGLRR"
2874. .3437
/gene="cut5b"
/locus_tag="RV3724B"
2874. .3437
/gene="cut5b"
/locus_tag="RV3724B"
/EC_number="3.1.1.-"
/function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE
STRUCTURE OF PLANT CUTICLE)."
/note="RV3724B, (MTV025.072), len: 187 aa. Probable cut5b,
truncated cutinase (EC 3.1.1.-), similar to C-terminal end
of others e.g. Q9XB09|RV2D-RV1758 PROTEIN (FRAGMENT) from
Mycobacterium bovis BCG (143 aa) FASTA scores: opt: 335,
E(): 3.4e-12, (53.25% identity in 92 aa overlap); O9K87
SERINE ESTERASE CUTINASE from Mycobacterium avium (220
aa), FASTA scores: opt: 251, E(): 2.5e-07, (44.05%
identity in 168 aa overlap). Also similar to proteins from
Mycobacterium tuberculosis e.g. O06793|RV1758|MTCY28.24
HYPOTHETICAL 17.9 KDA PROTEIN (174 aa), FASTA scores: opt:

```

641, E(): 2.5e-29, (57.25% identity in 166 aa overlap);
006319|RV3452|MYCY13B12.05 HYPOTHETICAL 23.1 KDA PROTEIN
(226 aa), FASTA scores: opt: 385, E(): 7.5e-15, (46.65%
identity in 165 aa overlap);
006318|CUT3_MYCTU|RV3451|MT557|MYCY13B12.04 PROBABLE
CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 307, E():
1.9e-10, (40.7% identity in 167 aa overlap);
010837|CUT1_MYCTU|RV1984c|MT2037|MYCY39.35 PROBABLE
CUTINASE PRECURSOR (217 aa), FASTA scores: opt: 261, E():
6.7e-08, (50.9% identity in 169 aa overlap); etc; and
U00015_11 from Mycobacterium lepra. 5'-end of gene is
RV3724A|CUT5A; frameshifting may occur near position
4169668. Tbpase score is 0.918."
/codon_start=1
/transl_table=11
/product="PROBABLE CUTINASE [SECOND PART] CUT5B"
/protein_id="CAE55633.1"
/db_xref="GI:38490373"
/translation="MAPGSHLVLAASEDCSTHCYSQVGAASLGVYAVNPASNDPAS
SDPKTVIDGIRDAGSHIQSMAMSCPTROYGYSQGAAGAAGVYTSVAVPAPVQCA
VPAPAPVAVNHVAIVTLFGAPSAQFLQYGAIPAIPLVQPKTLQCADGDSICSD
GNSPVAHGLYAVNGVVGAGNFAASRL"

Query Match 100.0%; Score 19; DB 1; Length 244800;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGCTTCGCTGGAGCTGG 19
|||||
Db 227659 TGGGCTTCGCTGGAGCTGG 227641

RESULT 12
BX248347/c 278492 bp DNA linear BCT 06-MAY-2004
LOCUS Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment
14/14.
ACCESSION BX248347 BX248333
VERSION BX248347.1 GI:31620456
KEYWORDS complete genome.
SOURCE Mycobacterium bovis AF2122/97
ORGANISM Mycobacterium bovis AF2122/97
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE 1
AUTHORS Garnier,T., Eiglmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
Pryor,M., Duchoy,S., Grondin,S., Lacroix,C., Monsenpe,C., Simon,S.,
Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,
Wheeler,P.R., Parkhill,J., Barrall,B.G., Cole,S.T., Gordon,S.V. and
Hewison,G.
TITLE The complete genome sequence of Mycobacterium bovis
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 278492)
Garnier,T.
REFERENCE Direct Submision
JOURNAL Submitted (24-MAR-2003) Garnier T., Unite de Genetique Molculaire
Bacterienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex
15, France. e-mail:fgarnier@pasteur.fr Submitted on behalf of the
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridgeshire CB10 1SA, UK. PT4 Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Molculaire Bacterienne, Institut Pasteur, 28
rue du Docteur Roux, 75724 Paris Cedex 15, France
LOCATION/Qualifiers

FEATURES
source
1..278492
/organism="Mycobacterium bovis AF2122/97"
/mol_type="genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"
complement(240..704)
gene

CDS

/locus_tag="MB3713c"
complement(240..704)
/locus_tag="MB3713c"
/note="MB3713c", len: 154 aa. Equivalent to RV3688c,
len: 154 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 154 aa overlap). Hypothetical protein,
similar to other bacterial hypothetical proteins e.g.
O9X341|SCHE6.02c from Streptomyces coelicolor (154 aa),
FASTA scores: opt: 425, E(): 3.4e-19, (46.1% identity in
154 aa overlap); O9WZFA|TM690 from Thermotoga maritima
(149 aa), FASTA scores: opt: 326, E(): 3.4e-13, (40.4%
identity in 151 aa overlap); O9PHU3|CU0573 from
Campylobacter jejuni (147 aa), FASTA scores: opt:290,
E(): 5.1e-11, (36.4% identity in 151 aa overlap); etc.
Also some similarity to upstream
O69654|RV3686c|MTV025.034c CONSERVED HYPOTHETICAL PROTEIN
from Mycobacterium tuberculosis.

gene

CDS

/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD95899.1"
/db_xref="GI:31620457"
/db_xref="UniProt/TREMBL:Q7TVX7"
/translation="MAELIKSQLRSDLTQAMKTQDKLRTATIRMLAAIQTEESGKQA
RELSDDEVITKVLARESRKRGRABAIYONGGEALAEHARARIIDEVLPPLTEGEL
ADVADTALAEVABEELGHRPSMKOMGLVKAATVIAAGKADGARLSAAVKERL"
704..2059
/locus_tag="MB3714"
704..2059
/locus_tag="MB3714"
/note="MB3714", len: 451 aa. Equivalent to RV3689, len:
451 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.8% identity in 451 aa overlap). Probable conserved
transmembrane protein, with proline rich N-terminus,
similar to O9KRW6|SC83.17 PUTATIVE INTEGRAL MEMBRANE
PROTEIN from Streptomyces coelicolor (462 aa) FASTA
scores: opt: 730, E(): 2.7e-21, (38.1% identity in 412 aa
overlap)."
/codon_start=1
/transl_table=11
/product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN"
/protein_id="CAD95900.1"
/db_xref="GI:31620458"
/db_xref="UniProt/TREMBL:Q7TVX6"
/translation="MHKRYAPORPKPDTEYIEKCTDRDQGDHDERQLRPVAMP
PGYVAPPPVAPAGVAPGYPVAPATPPGPGGAPGAPSPGPPGPGVAPAPAPG
CGPPPGYPPGYPPLAPAVKKGIIPLRLTSLDITNGAVCYIRAPKATLIGTAM
VVLTQIISVALGEPMTAFEDVITGEBDELFGAVGMSFGASLIVSMAGVLS
AAALVIVGAVFGSPITVGRAMAVRGLLAFGLALEAGVAVLGLAVVILSGVA
FMRVIGRLITVLVGVGNALIAAPFMVIGELIVAVTASDGSVTRVLGATLSAIGVT
IGQIVTAPFSGVVLVLTDRIRIABARDVLTQGLGAPGAPGAPVSTNLMLTRP
F"

gene

CDS

2086..2739
/locus_tag="MB3715"
2086..2739
/locus_tag="MB3715"
/note="MB3715", len: 217 aa. Equivalent to RV3690, len:
217 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 217 aa overlap). Probable conserved
membrane protein, similar to O9KY45|SC83.18 PUTATIVE
INTEGRAL MEMBRANE PROTEIN from Streptomyces coelicolor
(231 aa), FASTA scores: opt: 419, E(): 1.5e-19, (36.0%
identity in 211 aa overlap). Equivalent to AAK48159 from
Mycobacterium tuberculosis strain CDC1551 (233 aa) but
shorter 16 aa."
/codon_start=1
/transl_table=11
/product="PROBABLE CONSERVED MEMBRANE PROTEIN"
/protein_id="CAD95901.1"
/db_xref="GI:31620459"
/db_xref="UniProt/TREMBL:Q7TVX5"
/translation="WPSIDIDREAHQAQRELDKPIYKCSLTKELTDWIDELQYRI

complement(240..704)
/locus_tag="MB3713c"
/note="MB3713c", len: 154 aa. Equivalent to RV3688c,
len: 154 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 154 aa overlap). Hypothetical protein,
similar to other bacterial hypothetical proteins e.g.
O9X341|SCHE6.02c from Streptomyces coelicolor (154 aa),
FASTA scores: opt: 425, E(): 3.4e-19, (46.1% identity in
154 aa overlap); O9WZFA|TM690 from Thermotoga maritima
(149 aa), FASTA scores: opt: 326, E(): 3.4e-13, (40.4%
identity in 151 aa overlap); O9PHU3|CU0573 from
Campylobacter jejuni (147 aa), FASTA scores: opt:290,
E(): 5.1e-11, (36.4% identity in 151 aa overlap); etc.
Also some similarity to upstream
O69654|RV3686c|MTV025.034c CONSERVED HYPOTHETICAL PROTEIN
from Mycobacterium tuberculosis.

gene
LKGSSIGGMMFTLVLLILMLIAVTAALVAAARTRMRNGDYQLPAGQITPAOHR
STAEVAEGEWMMAAIRRLAIVAAARELEVTOMLPAAGRTACQASAEPLTAOER
LTQATATANDVTYGERPCTQGAHYOMIADLDHLSKSPAVVASVQHPAIPDSMAQR"
2865..3866
/locus_tag="Mb3716"
2865..3866
/locus_tag="Mb3716"
/note="Mb3716, -, len: 333 aa. Equivalent to RV3691, len: 333 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 333 aa overlap). Conserved hypothetical protein, similar to G9KW4|SCB33.19 PUTATIVE SECRETED PROTEIN from Streptomyces coelicolor (387 aa) FASTA scores: opt: 481, E(): 66-23, (36.6% identity in 358 aa overlap). Equivalent to AKR8160 from Mycobacterium tuberculosis strain CDC1551 (381 aa) but shorter 48 aa."
/codon_start=1
/transl_table=1
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD95902.1"
/db_xref="GI:31620460"
/db_xref="UniProt/TREMBL:Q7TVX4"
/translation="MAPASTSTGGHALATLLGNHGVEVAVVADSIADVEAARPSDILLVAQOYLVNDALDLRLAKAGSDLLVAPTERTALTPQRLIAASPSRNPCTIRAEARRGSVQMPESPDTYQATGDIVLTCGGLAVRRAEGRITVAGSEPNFGSLPLAGNAALANLNIAPRRLVAAPDHLIEGMSPPSSLDLIPENWYTIWQLVVLVAALWKGRIICPLVAEELPVVIRASFEVGEGRISRRADRADLRTALQRLPRGLGVAGAPAAVVVTTIQRSKADPPFAVYHLFGPATVDNDLLQLARALDIERQYHS"
3863..4939
/gene="moxR2"
/locus_tag="Mb3717"
3863..4939
/gene="moxR2"
/locus_tag="Mb3717"
/note="Mb3717, moxR2, len: 358 aa. Equivalent to RV3692, len: 358 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 358 aa overlap). Probable moxR, methanol dehydrogenase regulatory protein, highly similar (generally longer at N-terminus) to G9KW3|SCB33.20 PUTATIVE REGULATORY PROTEIN from Streptomyces coelicolor (329 aa), FASTA scores: opt: 1523, E(): 4.2e-74, (70.9% identity in 330 aa overlap); G9Z538|SC982.21c PUTATIVE REGULATORY PROTEIN from Streptomyces coelicolor (332 aa) FASTA scores: opt: 1008, E(): 1.1e-46, (50.8% identity in 313 aa overlap); Q9U267|MOXR-3|PAB0848 METHANOL DEHYDROGENASE REGULATORY PROTEIN from Pyrococcus abyssi (314 aa), FASTA scores: opt: 989, E(): 1.1e-45, (50.65% identity in 302 aa overlap); G9AANI|CC0566 MOXR PROTEIN from Callobacter crescentus (323 aa), FASTA scores: opt: 988, E(): 1.3e-45, (52.3% identity in 306 aa overlap); etc. Also similar to O53170|MYT007.26|MOXR|Fv1479 from Mycobacterium tuberculosis (377 aa); and O07332|AF002133_6|MOXR from Mycobacterium avium (309 aa). Also high similarity with several hypothetical bacterial proteins."
/codon_start=1
/transl_table=1
/product="PUTATIVE METHANOL DEHYDROGENASE TRANSCRIPTIONAL REGULATORY PROTEIN MOXR2"
/protein_id="CAD95903.1"
/db_xref="GI:31620461"
/db_xref="UniProt/TREMBL:Q7TVX3"
/translation="WTQASANDPAOPOTPPALRGPVPAAGAPTAAGSPGHYARAEASADALLARARVAKVAGVODGVSTSLYALICRGVHVLEGRVYAKTLIYANAKAALQLEKRVQTFPDLMPGDVGTSLYDARKLAEFFRRPPTVNTLLADEIRTPPKXTQAAHLAMBERDQVSVGEKPLPNPITVAATONLEBETSTQLPEAQRDLFLKLVNTLLPARDEILILDLHAHGDPDRLSAINPVAAGAEAAEAAVRAHVLVANEVYGLVIVDITGATRSPPDLQGLVSPRGATALTGTARSWAAMSGRDYVTPDVVKAMARPTLHRHWLRRPKELEGGATPDQGLVDISLAVPYVR"
5073..6395
/locus_tag="Mb3718"
5073..6395
/locus_tag="Mb3718"
/note="Mb3718, -, len: 440 aa. Equivalent to RV3693, len:

```

Query Match      100.0%; Score 19; DB 1; Length 278492;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 TGGGGTTCGCTGGAAGTGG 19
|||||||
Db      261352 TGGGGTTCGCTGGAAGTGG 261334

RESULT 13
CR847886/c
LOCUS
DEFINITION Mus musculus chromosome X clone W11-2134P12, WORKING DRAFT
SEQUENCE, 2 unordered pieces.
ACCESSION CR847886
VERSION CR847886.1 GI:53850224
KEYWORDS HTG; HTGS PHASE1; HTGS ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 43836)
Grafham, D.
Direct Submission
Submitted (05-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: FM2134P12
----- Summary Statistics
Assembly program: XGAP4, version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 43356 bases at least Q40
Consensus quality: 43507 bases at least Q30
Consensus quality: 43620 bases at least Q20
Insert size: 43736; sum-of-contigs
Insert size: 51468; 2.9% error; agarose-fp
Quality coverage: 6.21x in Q20 bases; sum-of-contigs Quality
coverage: 6.12x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

*      1      10908: contig of 10908 bp in length
*      11009      11008: gap of 100 bp
*      11009      43836: contig of 32828 bp in length.
FEATURES
SOURCE
1. 43836
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="X"
/clone="W11-2134P12"
/clone_11b="Whitehead_mouse_foemid"
1. 10908
/note="assembly_fragment:00146
fragment_chain:1"
11009. 43836
/note="assembly_fragment:00240
fragment_chain:1
clone_end:SP6
vector_side:right"
misc_feature
misc_feature
1. 10908
/note="assembly_fragment:00146
fragment_chain:1"
11009. 43836
/note="assembly_fragment:00240
fragment_chain:1
clone_end:SP6
vector_side:right"
ORIGIN
Query Match      91.6%; Score 17.4; DB 2; Length 43836;
Best Local Similarity 94.7%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy      1      TGGGGTTCGCTGCAAGTCG 19
DB      10688 TGGGGTTCGCTGCAAGTCG 10670

RESULT 14
AC115304      144026 bp      DNA      linear      ROD 05-NOV-2003
LOCUS      AC115304/C
DEFINITION      Mus musculus BAC clone RP24-156K1 from 15, complete sequence.
ACCESSION      AC115304
VERSION      AC115304.3 GI:21844659
KEYWORDS      HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerogamachi; Muridae; Murinae; Mus.
1 (bases 1 to 144026)
Mammalia; Eutheria; Rodentia; Sclerogamachi; Muridae; Murinae; Mus.
The sequence of Mus musculus BAC clone RP24-156K1
JOURNAL      Unpublished (2001)
REFERENCE      2 (bases 1 to 144026)
AUTHORS      Wilson, R.
TITLE      Sequencing of Mus musculus
JOURNAL      Unpublished (2001)
REFERENCE      3 (bases 1 to 144026)
AUTHORS      McPherson, J.D. and Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (16-MAR-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE      4 (bases 1 to 144026)
AUTHORS      McPherson, J.D. and Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE      5 (bases 1 to 144026)
AUTHORS      McPherson, J.D. and Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (16-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE      6 (bases 1 to 144026)
AUTHORS      Wilson, R.
TITLE      Direct Submission
JOURNAL      Submitted (05-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 16, 2002 this sequence version replaced gi:21040012.
-----
Center: Washington University Genome Sequencing Center
Center code: WUSC

```

```

Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics -----
Center project name: M_BB0156K01
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RP24-156K1 BAC library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
location/Qualifiers
1. 144026
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/map="15"
/clone="RP24-156K1"
/clone_11b="RP24-156K1"
1748. 1887
/rpc_family="Alu"
repeat_region
4935. 5023
/rpc_family="B2"
repeat_region
6425. 6472
/rpc_family="Alu"
repeat_region
6494. 6695
/rpc_family="B2"
repeat_region
7330. 7564
/rpc_family="CR1"
repeat_region
7621. 7734
/rpc_family="Alu"
repeat_region
7828. 7967
/rpc_family="B4"
repeat_region
8812. 8948
/rpc_family="B4"
repeat_region
10054. 10188
/rpc_family="Alu"
repeat_region
10289. 10472
/rpc_family="B2"
repeat_region
11445. 11563
/rpc_family="Alu"
repeat_region
11657. 11718
/rpc_family="Alu"
repeat_region
11667. 11743
/rpc_family="B4"
repeat_region
11933. 12035
/rpc_family="Alu"
repeat_region
12368. 12447
/rpc_family="L1"
repeat_region
12828. 12977

```

```
/rpt_family="B4"
repeat_region 1303T..13193
/rpt_family="B2"
repeat_region 1342T..13627
/rpt_family="MER1_type"
repeat_region 14279..14400
/rpt_family="B4"
repeat_region 14547..14679
/rpt_family="B2"
repeat_region 14716..15006
/rpt_family="MALR"
repeat_region 16277..16411
/rpt_family="Alu"
repeat_region 16416..16494
/rpt_family="ID"
repeat_region 17687..17927
/rpt_family="L1"
repeat_region 17929..18103
/rpt_family="B2"
repeat_region 18106..18223
/rpt_family="L1"
repeat_region 18293..18447
/rpt_family="MER2_type"
repeat_region 18448..18629
/rpt_family="B2"
complement(18551..18623)
/product="tRNA-Ser"
/note="Likely pseudogene (HMM Sc=39.59 / Sec struct
Sc=-11.21)"
repeat_region 18630..18678
/rpt_family="MER2_type"
repeat_region 18703..18828
/rpt_family="MER2_type"
repeat_region 19957..20261
/rpt_family="L1"
repeat_region 20364..20509
/rpt_family="Alu"
repeat_region 20841..21461
/rpt_family="L1"
repeat_region 21879..22025
/rpt_family="Alu"
repeat_region 22041..22148
/rpt_family="MALR"
repeat_region 22149..22501
/rpt_family="MALR"
repeat_region 22307..22533
/rpt_family="MALR"
repeat_region 22712..22988
/rpt_family="L1"
repeat_region 23969..24065
/rpt_family="MALR"
repeat_region 24058..24181
/rpt_family="MALR"
repeat_region 25028..25216
/rpt_family="MALR"
repeat_region 26526..26636
/rpt_family="MALR"
repeat_region 26687..26919
/rpt_family="ERV1"
repeat_region 26973..27576
/rpt_family="ERV1"
repeat_region 27966..28021
/rpt_family="ERV1"
repeat_region 28028..30206
/rpt_family="ERV1"
repeat_region 30666..30758
/rpt_family="MALR"
repeat_region 30955..31009
/rpt_family="MALR"
repeat_region 31260..31397
/rpt_family="B4"
repeat_region 31572..31710
/rpt_family="Alu"
```

```
repeat_region 32227..32297
/rpt_family="B4"
repeat_region 32306..32447
/rpt_family="Alu"
unstore 35042..35191
/note="Sequence derived from one plasmid subclone."
repeat_region 37307..37672
/rpt_family="ERV1"
repeat_region 37938..38097
/rpt_family="ERV1"
repeat_region 38359..38608
/rpt_family="B4"
repeat_region 38875..38986
```

Query Match 91.6%; Score 17.4; DB 10; Length 144026;
Best Local Similarity 94.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGTGGAAGTGG 19
|||||
Db 63490 TGGGGTTCGTGGAAGTGG 63472

RESULT 15
AC147366 175781 bp DNA linear ROD 13-JUL-2004
DEFINITION Mus musculus chromosome 15 clone RP23-386J3, complete sequence.
AC147366
VERSION AC147366.5 GI:50253696
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL Wilson, R.K.
The sequence of Mus musculus clone
unpublished
2 (bases 1 to 175781)

REFERENCE
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 175781)

REFERENCE
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 175781)

REFERENCE
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Jul 13, 2004 this sequence version replaced gi:47605149.

COMMENT ----- Genome Center -----

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
Project Information
Center project name: M_BA0386J03

FEATURES
Location/Qualifiers

1..175781
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/clone="RP23-386J3"

ORIGIN

Query Match 91.6%; Score 17.4; DB 10; Length 175781;

Best Local Similarity 94.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGGTTGGCTGGAGTGG 19
|||||
Db 8125 TGGGTTCTCTGGAGTGG 8143

Search completed: November 7, 2005, 06:42:07
Job time : 754.66 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:04:21 ; Search time 219.106 Seconds
(without alignments)
513.336 Million cell updates/sec

Title: US-10-777-131A-4
Perfect score: 19
Sequence: 1 tggggcttcgcgcgaagtcg 19

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

- 1: geneeqn1980s:*
- 2: geneeqn1990s:*
- 3: geneeqn2000s:*
- 4: geneeqn2001s:*
- 5: geneeqn2002as:*
- 6: geneeqn2002bs:*
- 7: geneeqn2002bs:*
- 8: geneeqn2003as:*
- 9: geneeqn2003bs:*
- 10: geneeqn2003cs:*
- 11: geneeqn2003ds:*
- 12: geneeqn2004as:*
- 13: geneeqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	100.0	19	ABZ79930	Abz79930 Mycobacte
2	19	100.0	1144	ABZ79953	Abz79953 Mycobacte
3	19	100.0	2406	ABZ79958	Abz79958 Mycobacte
4	19	100.0	2409	AAK34178	AAK34178 Mycobacte
5	19	100.0	2451	AAK34179	AAK34179 Mycobacte
6	19	100.0	2806	ABZ79957	Abz79957 Mycobacte
7	19	100.0	103765	AA199683_43	Continuation (44 o
8	19	100.0	110000	AA199682_43	Continuation (44 o
9	16.4	86.3	4428	ADN04195	Adn04195 Antispori
10	16.4	86.3	117328	ABD32886	Abd32886 Mouse can
11	15.8	83.2	292	AA125090	AA125090 Probe #15
12	15.8	83.2	292	ABA70776	ABA70776 Human toe
13	15.8	83.2	292	AA150949	AA150949 Probe #19
14	15.8	83.2	292	ABA37266	ABA37266 Probe #15
15	15.8	83.2	292	AAK44980	AAK44980 Human bon
16	15.8	83.2	292	AAK19029	AAK19029 Human bra
17	15.8	83.2	292	ABK44646	ABK44646 Human liv
18	15.8	83.2	292	ABK19225	ABK19225 Human gen
19	15.8	83.2	304	AA102296	AA102296 Human rep
20	15.8	83.2	445	AA190253	AA190253 Human pol

21	15.8	83.2	476	AA115873	AA115873 Probe #58
22	15.8	83.2	476	ABK58175	ABK58175 Human toe
23	15.8	83.2	476	AA137782	AA137782 Probe #64
24	15.8	83.2	476	ABA27360	ABA27360 Probe #58
25	15.8	83.2	476	AAK31915	AAK31915 Human bon
26	15.8	83.2	476	AAK06255	AAK06255 Human bra
27	15.8	83.2	476	ABK31605	ABK31605 Human liv
28	15.8	83.2	476	ABK06677	ABK06677 Human gen
29	15.8	83.2	738	ACB32274	ACB32274 Prokaryot
30	15.8	83.2	984	ADT4231	ADT4231 Bacterial
31	15.8	83.2	1814	AB199704	AB199704 Mouse isc
32	15.8	83.2	2213	ACC46652	ACC46652 Human dit
33	15.8	83.2	2511	ACA24500	ACA24500 Prokaryot
34	15.8	83.2	2725	ADB68895	ADB68895 C. neofor
35	15.8	83.2	3240	ADB62016	ADB62016 Human CDN
36	15.8	83.2	3268	AAK87591	AAK87591 DNA encod
37	15.8	83.2	5109	ACN42259	ACN42259 Human dia
38	15.8	83.2	5229	ACN42258	ACN42258 Human dia
39	15.8	83.2	5342	AAD00374	AAD00374 Rat smoot
40	15.8	83.2	5739	ABK70271	ABK70271 Human Jun
41	15.8	83.2	6327	AAK732301	AAK732301 Dermatomy
42	15.8	83.2	6417	ABL64410	ABL64410 Stomach C
43	15.8	83.2	6417	ABX04169	ABX04169 Human mRN
44	15.8	83.2	6417	ABK84696	ABK84696 Human CDN
45	15.8	83.2	6417	AAK54631	AAK54631 Human chr

ALIGNMENTS

RESULT 1	ABZ79930	standard; DNA, 19 BP.
ID	ABZ79930	
XX	ABZ79930;	
AC	19-MAY-2003 (first entry)	
XX		
DT		
XX		
DE	Mycobacterium tuberculosis Rv3908-2 PCR primer SEQ ID NO:4.	
XX		
KW	Mycobacterium tuberculosis; mutT2; alkA; ogt; Rv3908; mutY; Rv3909;	
KM	detection; multidrug resistance; multiple drug resistance; MDR;	
KX	infection; PCR primer; ss.	
OS	Mycobacterium tuberculosis.	
XX	Synthetic.	
PN	WO2003016562-A2.	
XX		
PD	27-FEB-2003.	
XX		
PF	14-AUG-2002; 2002WO-EP009679.	
XX		
PR	14-AUG-2001; 2001US-0311824P.	
XX		
PR	21-AUG-2001; 2001US-0313523P.	
XX		
PA	(INSP) INST PASTEUR.	
XX		
PI	Gicquel B;	
XX		
DR	WPI, 2003-256711/25.	
XX		
PT	Predicting the epidemic character of a Mycobacterium tuberculosis isolate	
XX	and/or the acquisition of multiple drug resistance (MDR) by the isolate	
PT	by detecting an alteration in the DNA repair system of the isolate.	
XX		
PS	Claim 32; Page 16; 83pp; English.	
XX		
CC	The present invention describes a method for predicting the epidemic	
CC	character of a Mycobacterium tuberculosis isolate and/or a selective	
CC	advantage to be maintained in the host and/or the acquisition of multiple	
CC	drug resistance (MDR) by the isolate comprising detecting an alteration	
CC	in the DNA repair system of the isolate. Also described: (1) detecting a	

CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a
CC polynucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)
CC an Escherichia coli strain containing the plasmid pMYC2501; and (5)
CC detecting in a patient infected by Mycobacterium tuberculosis a higher
CC risk of being unable to eliminate the bacillus or of developing MDR
CC tuberculosis. The method is useful for predicting the epidemic character
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to
CC be maintained in the host and/or the acquisition of MDR by the isolate.
CC The present sequence represents a PCR primer for M. tuberculosis RV3908,
CC which is used in the exemplification of the present invention
XX
SQ Sequence 19 BP; 2 A; 2 C; 10 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 8; Length 19;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGGGTTCGCTGGAAGTGG 19
Db 1 TGGGGTTCGCTGGAAGTGG 19

RESULT 2
ABZ79953/c 100.0%; Score 19; DB 8; Length 19;
ID ABZ79953 standard; DNA; 1144 BP.
XX
XX ABZ79953;
AC
XX
XX 19-MAY-2003 (first entry)
DT
XX
XX Mycobacterium tuberculosis RV3908 nucleotide sequence SEQ ID NO:20.
DE
XX
XX Mycobacterium tuberculosis; mutT2; alkA; ogt; RV3908; mutY; RV3909;
KM Mycobacterium tuberculosis; multidrug resistance; multiple drug resistance; MDR;
KM infection; gene; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX WO2003016562-A2.
PN
XX
XX 27-FEB-2003.
PD
XX
XX 14-AUG-2002; 2002WO-EP009679.
PF
XX
XX 14-AUG-2001; 2001US-0311824P.
PR 21-AUG-2001; 2001US-0313523P.
XX
XX (INSP) INST PASTEUR.
PA
XX
XX Glcquel B;
PI
XX
XX WPI; 2003-256711/25.
DR
XX
XX
PT Predicting the epidemic character of a Mycobacterium tuberculosis isolate
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate
PT by detecting an alteration in the DNA repair system of the isolate.
PS
XX
XX Disclosure; Fig 5A; 83pp; English.
XX
XX The present invention describes a method for predicting the epidemic
XX character of a Mycobacterium tuberculosis isolate and/or a selective
XX advantage to be maintained in the host and/or the acquisition of multiple
XX drug resistance (MDR) by the isolate comprising detecting an alteration
XX in the DNA repair system of the isolate. Also described: (1) detecting a
XX Mycobacterium tuberculosis strain with a MDR phenotype; (2) a
XX polynucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)
XX an Escherichia coli strain containing the plasmid pMYC2501; and (5)
XX detecting in a patient infected by Mycobacterium tuberculosis a higher
XX risk of being unable to eliminate the bacillus or of developing MDR
XX tuberculosis. The method is useful for predicting the epidemic character
XX of a Mycobacterium tuberculosis isolate and/or a selective advantage to
XX be maintained in the host and/or the acquisition of MDR by the isolate.
XX The present sequence represents a M. tuberculosis RV3908 nucleotide

CC sequence, which is used in the exemplification of the present invention
XX
SQ Sequence 1144 BP; 194 A; 405 C; 379 G; 166 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 8; Length 1144;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGGGTTCGCTGGAAGTGG 19
Db 1143 TGGGGTTCGCTGGAAGTGG 1125

RESULT 3
ABZ79958/c 100.0%; Score 19; DB 8; Length 2406;
ID ABZ79958 standard; DNA; 2406 BP.
XX
XX ABZ79958;
AC
XX
XX 19-MAY-2003 (first entry)
DT
XX
XX Mycobacterium tuberculosis RV3909 ORF sequence SEQ ID NO:32.
DE
XX
XX Mycobacterium tuberculosis; mutT2; alkA; ogt; RV3908; mutY; RV3909;
KM Mycobacterium tuberculosis; multidrug resistance; multiple drug resistance; MDR;
KM infection; gene; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX WO2003016562-A2.
PN
XX
XX 27-FEB-2003.
PD
XX
XX 14-AUG-2002; 2002WO-EP009679.
PF
XX
XX 14-AUG-2001; 2001US-0311824P.
PR 21-AUG-2001; 2001US-0313523P.
XX
XX (INSP) INST PASTEUR.
PA
XX
XX Glcquel B;
PI
XX
XX WPI; 2003-256711/25.
DR
XX
XX
PT Predicting the epidemic character of a Mycobacterium tuberculosis isolate
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate
PT by detecting an alteration in the DNA repair system of the isolate.
PS
XX
XX Claim 45; Fig 7C; 83pp; English.
XX
XX The present invention describes a method for predicting the epidemic
XX character of a Mycobacterium tuberculosis isolate and/or a selective
XX advantage to be maintained in the host and/or the acquisition of multiple
XX drug resistance (MDR) by the isolate comprising detecting an alteration
XX in the DNA repair system of the isolate. Also described: (1) detecting a
XX Mycobacterium tuberculosis strain with a MDR phenotype; (2) a
XX polynucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)
XX an Escherichia coli strain containing the plasmid pMYC2501; and (5)
XX detecting in a patient infected by Mycobacterium tuberculosis a higher
XX risk of being unable to eliminate the bacillus or of developing MDR
XX tuberculosis. The method is useful for predicting the epidemic character
XX of a Mycobacterium tuberculosis isolate and/or a selective advantage to
XX be maintained in the host and/or the acquisition of MDR by the isolate.
XX The present sequence represents a M. tuberculosis RV3909 open reading
XX frame (ORF) sequence, which is used in the exemplification of the present
XX invention
XX
SQ Sequence 2406 BP; 369 A; 899 C; 760 G; 378 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 8; Length 2406;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTCGCTGGAAGTGG 19
|||||
Db 200 TGGGGTTCGCTGGAAGTGG 182

RESULT 4

AAK34178/c

ID AAK34178 standard; DNA; 2409 BP.

XX AAK34178;

DT 06-JUL-1999 (first entry)

DE Mycobacterium species nucleic acid sequence 36D.

KM Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
hybridisation; detection; vaccine; immunisation; infection; ss.

OS Mycobacterium sp.

PN WO9909186-A2.

PD 25-FEB-1999.

PF 14-AUG-1998; 98WO-FR001813.

PR 14-AUG-1997; 97FR-00010404.

PR 11-SEP-1997; 97FR-00011325.

XX (INSP) INST PASTEUR.

XX Gicquel B, Portnoie D, Lim E, Pellicic V, Guigueno A;

PI Goguet De La Salmoniere Y;

XX WPI; 1999-181045/15.

DR P-PSDB; AAY04925.

PT Mycobacterial DNA vectors containing reporter constructs - for

PT Identifying coding or promoter sequences involved in infection-associated

PS protein expression.

PS Claim 22; Fig 36D; 309pp; French.

CC Sequences AAK34001-X34252 represent nucleic acids encoding secreted
proteins from various Mycobacterium species microorganisms. The
CC nucleotide sequences can be used as primers and probes for methods for
detecting and identifying mycobacteria, especially belonging to the M.
CC tuberculosis complex. The encoded proteins can be used in vaccines for
immunisation against a bacterial or viral infection

SQ Sequence 2409 BP; 370 A; 899 C; 761 G; 379 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 2; Length 2409;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTCGCTGGAAGTGG 19
|||||
Db 200 TGGGGTTCGCTGGAAGTGG 182

RESULT 5

AAK34179/c

ID AAK34179 standard; DNA; 2451 BP.

XX AAK34179;

DT 06-JUL-1999 (first entry)

DE Mycobacterium species nucleic acid sequence 36F.

KM Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
hybridisation; detection; vaccine; immunisation; infection; ss.

XX Mycobacterium sp.
OS
XX
XX WO9909186-A2.
PN
XX
XX
XX
XX

PD 25-FEB-1999.

PF 14-AUG-1998; 98WO-FR001813.

PR 14-AUG-1997; 97FR-00010404.

PR 11-SEP-1997; 97FR-00011325.

XX (INSP) INST PASTEUR.

XX Gicquel B, Portnoie D, Lim E, Pellicic V, Guigueno A;

PI Goguet De La Salmoniere Y;

XX WPI; 1999-181045/15.

DR P-PSDB; AAY04926.

PT Mycobacterial DNA vectors containing reporter constructs - for

PT Identifying coding or promoter sequences involved in infection-associated

PS protein expression.

PS Claim 22; Fig 36F; 309pp; French.

CC Sequences AAK34001-X34252 represent nucleic acids encoding secreted
proteins from various Mycobacterium species microorganisms. The
CC nucleotide sequences can be used as primers and probes for methods for
detecting and identifying mycobacteria, especially belonging to the M.
CC tuberculosis complex. The encoded proteins can be used in vaccines for
immunisation against a bacterial or viral infection

SQ Sequence 2451 BP; 377 A; 915 C; 776 G; 383 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 2; Length 2451;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTCGCTGGAAGTGG 19
|||||
Db 242 TGGGGTTCGCTGGAAGTGG 224

RESULT 6

ABZ79957/c

ID ABZ79957 standard; DNA; 2806 BP.

XX ABZ79957;

DT 19-MAY-2003 (first entry)

DE Mycobacterium tuberculosis Rv3909 nucleotide sequence SEQ ID NO:22.

KM Mycobacterium tuberculosis; mutT2; alKA; oqT; Rv3908; mutY; Rv3909;

KM infection; multidrug resistance; multiple drug resistance; MDR;

XX infection; gene; ds.

OS Mycobacterium tuberculosis.

PN WO2003016562-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-BP009679.

PR 14-AUG-2001; 2001US-0311824P.

PR 21-AUG-2001; 2001US-0313523P.

XX (INSP) INST PASTEUR.

XX Gicquel B;

DR WP1; 2003-256711/25.
XX Predicting the epidemic character of a Mycobacterium tuberculosis isolate
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate
PT by detecting an alteration in the DNA repair system of the isolate.
PS Disclosure; Fig 7A-B; 83pp; English.
XX The present invention describes a method for predicting the epidemic
CC character of a Mycobacterium tuberculosis isolate and/or a selective
CC advantage to be maintained in the host and/or the acquisition of multiple
CC drug resistance (MDR) by the isolate comprising detecting an alteration
CC in the DNA repair system of the isolate. Also described: (1) detecting a
CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a
CC polynucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)
CC an *Escherichia coli* strain containing the plasmid pMYC2501; and (5)
CC detecting in a patient infected by Mycobacterium tuberculosis a higher
CC risk of being unable to eliminate the bacillus or of developing MDR
CC tuberculosis. The method is useful for predicting the epidemic character
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to
CC be maintained in the host and/or the acquisition of MDR by the isolate.
CC The present sequence represents a M. tuberculosis Rv3909 nucleotide
CC sequence, which is used in the exemplification of the present invention
CC
SQ Sequence 2806 BP; 433 A; 1048 C; 886 G; 439 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 8; Length 2806;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TGGGTTGCTGGAAGTG 19
Db 400 TGGGTTGCTGGAAGTG 382

RESULT 7

AA199683_43/c
Continuation (44 of 44) of AA199683 from base 4300001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AA199683 Accession AA199683

WP	Fragment Name	Begin	End
WP	AA199683_00	1	110000
WP	AA199683_01	100001	210000
WP	AA199683_02	200001	310000
WP	AA199683_03	300001	410000
WP	AA199683_04	400001	510000
WP	AA199683_05	500001	610000
WP	AA199683_06	600001	710000
WP	AA199683_07	700001	810000
WP	AA199683_08	800001	910000
WP	AA199683_09	900001	1010000
WP	AA199683_10	1000001	1110000
WP	AA199683_11	1100001	1210000
WP	AA199683_12	1200001	1310000
WP	AA199683_13	1300001	1410000
WP	AA199683_14	1400001	1510000
WP	AA199683_15	1500001	1610000
WP	AA199683_16	1600001	1710000
WP	AA199683_17	1700001	1810000
WP	AA199683_18	1800001	1910000
WP	AA199683_19	1900001	2010000
WP	AA199683_20	2000001	2110000
WP	AA199683_21	2100001	2210000
WP	AA199683_22	2200001	2310000
WP	AA199683_23	2300001	2410000
WP	AA199683_24	2400001	2510000
WP	AA199683_25	2500001	2610000
WP	AA199683_26	2600001	2710000
WP	AA199683_27	2700001	2810000
WP	AA199683_28	2800001	2910000
WP	AA199683_29	2900001	3010000
WP	AA199683_30	3000001	3110000
WP	AA199683_31	3100001	3210000
WP	AA199683_32	3200001	3310000

WP	AA199683_33	3300001	3410000
WP	AA199683_34	3400001	3510000
WP	AA199683_35	3500001	3610000
WP	AA199683_36	3600001	3710000
WP	AA199683_37	3700001	3810000
WP	AA199683_38	3800001	3910000
WP	AA199683_39	3900001	4010000
WP	AA199683_40	4000001	4110000
WP	AA199683_41	4100001	4210000
WP	AA199683_42	4200001	4310000
WP	AA199683_43	4300001	4403765

Query Match 100.0%; Score 19; DB 4; Length 103765;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TGGGTTGCTGGAAGTG 19
Db 86625 TGGGTTGCTGGAAGTG 86607

RESULT 8

AA199682_43/c
Continuation (44 of 45) of AA199682 from base 4300001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AA199682 Accession AA199682

WP	Fragment Name	Begin	End
WP	AA199682_00	1	110000
WP	AA199682_01	100001	210000
WP	AA199682_02	200001	310000
WP	AA199682_03	300001	410000
WP	AA199682_04	400001	510000
WP	AA199682_05	500001	610000
WP	AA199682_06	600001	710000
WP	AA199682_07	700001	810000
WP	AA199682_08	800001	910000
WP	AA199682_09	900001	1010000
WP	AA199682_10	1000001	1110000
WP	AA199682_11	1100001	1210000
WP	AA199682_12	1200001	1310000
WP	AA199682_13	1300001	1410000
WP	AA199682_14	1400001	1510000
WP	AA199682_15	1500001	1610000
WP	AA199682_16	1600001	1710000
WP	AA199682_17	1700001	1810000
WP	AA199682_18	1800001	1910000
WP	AA199682_19	1900001	2010000
WP	AA199682_20	2000001	2110000
WP	AA199682_21	2100001	2210000
WP	AA199682_22	2200001	2310000
WP	AA199682_23	2300001	2410000
WP	AA199682_24	2400001	2510000
WP	AA199682_25	2500001	2610000
WP	AA199682_26	2600001	2710000
WP	AA199682_27	2700001	2810000
WP	AA199682_28	2800001	2910000
WP	AA199682_29	2900001	3010000
WP	AA199682_30	3000001	3110000
WP	AA199682_31	3100001	3210000
WP	AA199682_32	3200001	3310000
WP	AA199682_33	3300001	3410000
WP	AA199682_34	3400001	3510000
WP	AA199682_35	3500001	3610000
WP	AA199682_36	3600001	3710000
WP	AA199682_37	3700001	3810000
WP	AA199682_38	3800001	3910000
WP	AA199682_39	3900001	4010000
WP	AA199682_40	4000001	4110000
WP	AA199682_41	4100001	4210000
WP	AA199682_42	4200001	4310000
WP	AA199682_43	4300001	4410000
WP	AA199682_44	4400001	4411529

Query Match 100.0%; Score 19; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTCGCTGGAAGTGG 19
|||||
DB 94388 TGGGGTTCGCTGGAAGTGG 94370

RESULT 9

ADN04199
ID ADN04199 standard; cDNA; 4428 BP.
XX
XX
AC ADN04199;
XX
XX
DT 01-JUL-2004 (first entry)
XX
DE Antipsoriatic cDNA sequence #299.
XX
XX ds; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
XX Homo sapiens.
XX
XX WO2004028479-A2.
XX
XX PD 08-APR-2004.
XX
XX PF 25-SEP-2003; 2003WO-US030907.
XX
XX PR 25-SEP-2002; 2002US-0414006P.
XX
XX (GENTH) GENENTECH INC.
XX
XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
PI Wu TD;
XX
XX WPI; 2004-305105/28.
XX
XX DR P-PSDB; ADN04200.
XX
XX PT New PRO nucleic acid or polypeptide, useful for preparing a
PT pharmaceutical composition for diagnosing or treating psoriasis in a
XX mammal.
XX
XX PS Claim 1; SEQ ID NO 593; 3069pp; English.
XX
XX CC The invention relates to novel polynucleotide and polypeptides for
XX treating psoriasis or a sequence having at least 80% identity to the
XX above sequences. The nucleic acid is useful for preparing a composition
XX for diagnosing or treating psoriasis in a mammal. This sequence
XX corresponds to one of the polynucleotides of the invention.
XX
XX SQ Sequence 4428 BP; 1174 A; 1006 C; 1009 G; 1239 T; 0 U; 0 Other;
XX
XX Query Match 86.3%; Score 16.4; DB 12; Length 4428;
XX Best Local Similarity 94.4%; Pred. No. 3e+02;
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 TGGGGTTCGCTGGAAGTGG 18
XX |||||
XX DB 753 TGGGGTTCGCTGGAAGTGG 770

RESULT 10

ABD32886/C
ID ABD32886 standard; DNA; 117328 BP.
XX
XX ABD32886;
XX
XX DT 18-NOV-2004 (first entry)
XX
XX DE Mouse cancer-associated genomic DNA MD18-013.
XX
XX KW Mouse; ds; cancer-associated protein; gene; cytosolic; cancer;
KW leukaemia; lymphoma; CAP.

XX Mus musculus.
OS
XX
XX WO2004074320-A2.
XX
XX PD 02-SEP-2004.
XX
XX PF 17-FEB-2004; 2004WO-US004730.
XX
XX PR 14-FEB-2003; 2003US-00367094.
XX
XX PR 14-MAR-2003; 2003US-00388838.
XX
XX PR 15-APR-2003; 2003US-00417375.
XX
XX PR 13-JUN-2003; 2003US-00461862.
XX
XX PR 15-SEP-2003; 2003US-00663431.
XX
XX PR 15-DEC-2003; 2003US-00737318.
XX
XX PA (SAGR-) SAGRES DISCOVERY INC.
XX
XX PI Morris DW, Morris DW, Malandro MS;
XX
XX DR WPI; 2004-652914/63.
XX
XX PT New isolated cancer-associated polynucleotides and polypeptides useful
XX for diagnosing, preventing or treating cancers, especially lymphoma and
XX leukemia, or in screening for agents that modulate cancer.
XX
XX PS disclosure; seqid 577; 310pp; English.

XX The invention relates to an isolated nucleic acid comprising at least 10
XX contiguous nucleotides of any of the 233 polynucleotide sequences given
XX in the specification, or its complement. The nucleic acid encodes cancer-
XX associated proteins. Also included are an expression vector comprising
XX the isolated nucleic acid cited above, a host cell comprising the above
XX recombinant nucleic acid or expression vector, a microarray for detecting
XX a cancer-associated (CA) nucleic acid comprising at least one probe
XX comprising at least 10 contiguous nucleotides of any of the above-
XX mentioned nucleotide sequences, an isolated polypeptide (encoded within
XX an open reading frame of a CA sequence selected from any of the 95
XX polynucleotide sequences as mentioned in the specification, or its
XX complement), an isolated antibody, (or its antigen binding fragment) that
XX binds to the above polypeptide, a hydridoma that produces the above
XX monoclonal antibody, a pharmaceutical composition comprising the above
XX antibody and a pharmaceutical excipient, a kit for detecting cancer
XX cells (comprising the antibody cited above, methods for diagnosing cancer
XX or for detecting the presence or absence of cancer cells in an
XX individual, a method for inhibiting growth of cancer cells in an
XX individual, a method for delivering a therapeutic agent to cancer cells
XX in an individual, an electronic library comprising the above
XX polynucleotide or polypeptide (or their fragments), methods of screening
XX for anticancer activity or for a bioactive agent capable of modulating
XX the activity of a CA protein (CAP), methods for detecting cancer
XX associated with expression of a polypeptide in a test cell sample, a
XX method for treating cancers and a method for inhibiting the expression of
XX CA gene in a cell. The composition and methods are useful for detecting,
XX diagnosing, preventing and treating cancers, especially lymphoma and
XX leukaemia. These may also be used in screening for agents that modulate
XX cancer. The present sequence is a mouse CAP genomic sequence. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 117328 BP; 31499 A; 25773 C; 25903 G; 33390 T; 0 U; 763 Other;

Query Match 86.3%; Score 16.4; DB 13; Length 117328;
Best Local Similarity 94.4%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGGTTGCTGGAAGTGG 19
|||||
DB 14283 GGGGTTGCTGGAAGTGG 14266

RESULT 11

```

AAI25090
ID AAI25090 standard; DNA; 292 BP.
XX
AC AAI25090;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #15023 for gene expression analysis in human cervical cell sample.
XX
KM Probe; human; microarray; gene expression; cervical epithelial cell;
XX
KM cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN MO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000670.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX
PT gene expression in human cervical epithelial cells.
XX
PS Claim 25; SEQ ID NO 15023; 487bp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
XX
CC (SENPs). The present sequence is one such probe. The SENPs are derived
XX
CC from human HeLa cells. The SENPs can be used to produce a single exon
XX
CC microarray, which can be used for measuring human gene expression in a
XX
CC sample derived from human cervical epithelial cells. By measuring gene
XX
CC expression, the probes are therefore useful in grading and/or staging of
XX
CC diseases of the cervix, notably cervical cancer. Note: The sequence data
XX
CC for this patent did not form part of the printed specification, but was
XX
CC obtained in electronic format directly from WIPO at
XX
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;
XX
Query Match 83.2%; Score 15.8; DB 4; Length 292;
XX
Best Local Similarity 89.5%; Pred. No. 5.1e+02;
XX
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 TGGGGTTCCTGGAAGTG 19
XX
DB 30 TGGGGATCTCTGGAAGTG 48
XX
RESULT 12
ABA70776
ID ABA70776 standard; DNA; 292 BP.
XX
AC ABA70776;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #19081.
XX
KM Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX

```

```

OS Homo sapiens.
XX
PN MO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000669.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX
PT gene expression in human fetal liver.
XX
PS Claim 4; SEQ ID NO 19081; 639pp + Sequence Listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for measuring
XX
CC human gene expression in a sample derived from human foetal liver. The
XX
CC single exon nucleic acid probes may be used for predicting, measuring and
XX
CC displaying gene expression in samples derived from human fetal liver. The
XX
CC present sequence is a single exon nucleic acid probe of the invention.
XX
CC Note: The sequence data for this patent did not form part of the printed
XX
CC specification, but was obtained in electronic format directly from WIPO
XX
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;
XX
Query Match 83.2%; Score 15.8; DB 4; Length 292;
XX
Best Local Similarity 89.5%; Pred. No. 5.1e+02;
XX
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 TGGGGTTCCTGGAAGTG 19
XX
DB 30 TGGGGATCTCTGGAAGTG 48
XX
RESULT 13
AAI50949
ID AAI50949 standard; DNA; 292 BP.
XX
AC AAI50949;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #19635 used to measure gene expression in human placenta sample.
XX
KM Probe; microarray; human; placenta; antenatal diagnosis;
XX
KM genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN MO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000663.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX

```

PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human placenta.
XX
XX
PS Claim 25; SEQ ID NO 19635; 654BP; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders
XX
SQ Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 4; Length 292;
Best Local Similarity 89.5%; Pred. No. 5.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGGGTTTCGCTGGAAGTGG 19
DB 30 TGGGGATCTCTGGAAGTGG 48

RESULT 14
ABA37266
ID ABA37266 standard; DNA; 292 BP.
XX
AC ABA37266;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #15732 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000666.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts.
XX
PS Claim 4; SEQ ID NO 15732; 530BP; English.
XX
CC The present invention relates to single exon nucleic acid probes for

CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at http://wipo.int/pat/published_pct_sequences
XX
SQ Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 4; Length 292;
Best Local Similarity 89.5%; Pred. No. 5.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGGGTTTCGCTGGAAGTGG 19
DB 30 TGGGGATCTCTGGAAGTGG 48

RESULT 15
AAK4980
ID AAK4980 standard; DNA; 292 BP.
XX
AC AAK4980;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 19537.
XX
DE Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000668.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO 19537; 658BP + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention
XX
SQ Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 4; Length 292;
Best Local Similarity 89.5%; Pred. No. 5.1e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TGGGTTGCTGGAGTGG 19
 |||||
 Db 30 TGGGATCTTGGAGTGG 48
 |||||

Search completed: November 7, 2005, 05:40:33
 Job time : 233.106 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: November 6, 2005, 19:19:37 ; Search time 64.4787 Seconds
(without alignments)
482.163 Million cell updates/sec

Title: US-10-777-131A-4

Perfect score: 19
Sequence: 1 tggggcttcgctgaagtgc 19

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A COMB. seq: *
2: /cgn2_6/prodata/1/ina/5B COMB. seq: *
3: /cgn2_6/prodata/1/ina/6A COMB. seq: *
4: /cgn2_6/prodata/1/ina/6B COMB. seq: *
5: /cgn2_6/prodata/1/ina/PC/TUS COMB. seq: *
6: /cgn2_6/prodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19	100.0	4403765	3 US-09-103-840A-2	Sequence 2, Appli
C 2	19	100.0	4411529	3 US-09-103-840A-1	Sequence 1, Appli
C 3	15.8	83.2	2678	4 US-09-807-757C-8	Sequence 8, Appli
C 4	15.8	83.2	5342	4 US-09-807-757C-1	Sequence 1, Appli
C 5	15.8	83.2	6328	3 US-08-913-832A-1	Sequence 1, Appli
C 6	15.8	83.2	6328	3 US-09-249-181A-1	Sequence 1, Appli
C 7	15.8	83.2	6328	4 US-09-158-707-1	Sequence 325, App
C 8	15.8	83.2	6475	4 US-09-620-312D-325	Sequence 13910, A
C 9	15.8	83.2	40887	4 US-09-949-016-13910	Sequence 13912, A
C 10	15.8	83.2	40887	4 US-09-949-016-13911	Sequence 13913, A
C 11	15.8	83.2	40887	4 US-09-949-016-13912	Sequence 13913, A
C 12	15.8	83.2	40887	4 US-09-949-016-13913	Sequence 14409, A
C 13	15.8	83.2	40887	4 US-09-949-016-14409	Sequence 14410, A
C 14	15.8	83.2	40887	4 US-09-949-016-14410	Sequence 14411, A
C 15	15.8	83.2	40887	4 US-09-949-016-14411	Sequence 14412, A
C 16	15.8	83.2	40887	4 US-09-949-016-14412	Sequence 15578, A
C 17	15.8	83.2	16291	4 US-09-949-016-15578	Sequence 15578, A
C 18	15.4	81.1	570	4 US-09-621-976-313	Sequence 313, App
C 19	15.4	81.1	872	4 US-09-786-240-23	Sequence 23, Appli
C 20	15.4	81.1	1007	4 US-09-949-016-917	Sequence 917, App
C 21	15.4	81.1	1007	4 US-09-949-016-3973	Sequence 3973, Ap
C 22	15.4	81.1	4773	4 US-09-949-016-2552	Sequence 2552, Ap
C 23	15.4	81.1	6415	4 US-09-949-016-1015	Sequence 1015, Ap
C 24	15.4	81.1	7388	4 US-09-949-016-15801	Sequence 15901, A
C 25	15.4	81.1	80858	4 US-09-949-016-12659	Sequence 12659, A
C 26	15.4	81.1	80859	4 US-09-949-016-15715	Sequence 15715, A
C 27	15	78.9	965	4 US-09-270-767-9078	Sequence 9078, Ap

28	15	78.9	965	4 US-09-270-767-24360	Sequence 24360, A
C 29	14.8	77.9	567	4 US-09-893-737-15	Sequence 15, Appli
C 30	14.8	77.9	601	4 US-09-949-016-71607	Sequence 71607, A
C 31	14.8	77.9	786	4 US-09-902-540-3126	Sequence 3126, Ap
C 32	14.8	77.9	817	4 US-09-809-545A-7	Sequence 7, Appli
C 33	14.8	77.9	1443	4 US-09-252-991A-3155	Sequence 3155, Ap
C 34	14.8	77.9	1929	4 US-09-252-991A-2815	Sequence 2815, Ap
C 35	14.8	77.9	9775	3 US-08-977-171-1	Sequence 1, Appli
C 36	14.8	77.9	15268	4 US-09-802-540-1142	Sequence 1142, Ap
C 37	14.8	77.9	31300	4 US-09-949-016-16967	Sequence 16967, A
C 38	14.8	77.9	32065	4 US-09-949-016-12136	Sequence 12136, A
C 39	14.8	77.9	32066	4 US-09-949-016-12136	Sequence 13268, A
C 40	14.8	77.9	38371	4 US-09-949-016-12061	Sequence 12061, A
C 41	14.8	77.9	38371	4 US-09-949-016-12488	Sequence 12488, A
C 42	14.8	77.9	38371	4 US-09-949-016-15596	Sequence 15596, A
C 43	14.8	77.9	38371	4 US-09-949-016-15597	Sequence 15597, A
C 44	14.8	77.9	46885	4 US-09-949-016-13848	Sequence 13848, A
C 45	14.8	77.9	50797	4 US-09-949-016-16346	Sequence 16346, A

ALIGNMENTS

```
RESULT 1
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match          100.0%; Score 19; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGGGTTCGCTGAAGTGC 19
Db      4386625 TGGGGTTCGCTGAAGTGC 4386607

RESULT 2
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
```

```
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 4411529
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match          100.0%; Score 19; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGGGTTGCTGGAAGTGG 19
DB      4394388 TGGGGTTGCTGGAAGTGG 4394370

RESULT 3
US-09-807-757C-8/C
/ Sequence 8, Application US/09807757C
/ Patent No. 6825035
/ GENERAL INFORMATION:
/ APPLICANT: Owens, Gary K.
/ APPLICANT: Mack, Christopher
/ APPLICANT: Blank, Randall
/ APPLICANT: University of Virginia Patent Foundation
/ TITLE OF INVENTION: Compositions and Methods for Modulating Expression
/ FILE REFERENCE: 021258-000500US
/ CURRENT APPLICATION NUMBER: US/09/807,757C
/ PRIOR FILING DATE: 2001-04-17
/ PRIOR APPLICATION NUMBER: US 60/105,330
/ PRIOR FILING DATE: 1998-10-23
/ PRIOR APPLICATION NUMBER: WO PCT/US99/24972
/ NUMBER OF SEQ ID NOS: 32
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 8
/ LENGTH: 2678
/ TYPE: DNA
/ ORGANISM: Rattus sp.
/ FEATURE:
/ OTHER INFORMATION: smooth muscle alpha-actin (SM alpha-A) gene
US-09-807-757C-8

Query Match          83.2%; Score 15.8; DB 4; Length 2678;
Best Local Similarity 89.5%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGGGTTGCTGGAAGTGG 19
DB      1846 TGGGGTTCTTGGAAAGTGG 1828

RESULT 4
US-09-807-757C-1/C
/ Sequence 1, Application US/09807757C
/ Patent No. 6825035
/ GENERAL INFORMATION:
/ APPLICANT: Owens, Gary K.
/ APPLICANT: Mack, Christopher
/ APPLICANT: Blank, Randall
/ APPLICANT: University of Virginia Patent Foundation
/ TITLE OF INVENTION: Compositions and Methods for Modulating Expression
/ FILE REFERENCE: 021258-000500US
/ CURRENT APPLICATION NUMBER: US/09/807,757C
/ PRIOR FILING DATE: 2001-04-17
/ PRIOR APPLICATION NUMBER: US 60/105,330
/ PRIOR FILING DATE: 1998-10-23
/ PRIOR APPLICATION NUMBER: WO PCT/US99/24972
/ PRIOR FILING DATE: 1999-10-22
```

```
/ NUMBER OF SEQ ID NOS: 32
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 5342
/ TYPE: DNA
/ ORGANISM: Rattus sp.
/ FEATURE:
/ OTHER INFORMATION: smooth muscle alpha-actin (SM alpha-A) gene
/ OTHER INFORMATION: regulatory region 5' promoter and intron
US-09-807-757C-1

Query Match          83.2%; Score 15.8; DB 4; Length 5342;
Best Local Similarity 89.5%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGGGTTGCTGGAAGTGG 19
DB      4452 TGGGGTTCTTGGAAAGTGG 4434

RESULT 5
US-08-913-832A-1/C
/ Sequence 1, Application US/08913832A
/ Patent No. 6329517
/ GENERAL INFORMATION:
/ APPLICANT: Seelig, Hans Peter
/ APPLICANT: Renz, Manfred
/ TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
/ FILE REFERENCE: 8484-0030-999
/ CURRENT APPLICATION NUMBER: US/08/913,832A
/ PRIOR FILING DATE: 1998-01-12
/ PRIOR APPLICATION NUMBER: PCT/DE96/00444
/ PRIOR FILING DATE: 1996-03-08
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 6328
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)...(5736)
US-08-913-832A-1

Query Match          83.2%; Score 15.8; DB 3; Length 6328;
Best Local Similarity 89.5%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGGGTTGCTGGAAGTGG 19
DB      1453 TGGGATCTCTGGAAGTGG 1435

RESULT 6
US-09-249-181A-1/C
/ Sequence 1, Application US/09249181A
/ Patent No. 6440679
/ GENERAL INFORMATION:
/ APPLICANT: Seelig, Hans Peter
/ APPLICANT: Renz, Manfred
/ TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
/ FILE REFERENCE: 8484-0059-999
/ CURRENT APPLICATION NUMBER: US/09/249,181A
/ PRIOR FILING DATE: 1999-02-12
/ PRIOR APPLICATION NUMBER: US 08/913,832
/ PRIOR FILING DATE: 1998-01-12
/ PRIOR APPLICATION NUMBER: PCT/DE96/00444
/ PRIOR FILING DATE: 1996-03-08
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 6328
```


TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1) ... (5736)
US-09-249-181A-1

Query Match
Best Local Similarity 83.2%; Score 15.8; DB 3; Length 6328;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19
Db 1453 TGGGGATCTCTGGAAGTGG 1435

RESULT 7
US-09-158-707-1/c
Sequence 1, Application US/09158707
Patent No. 6500923
GENERAL INFORMATION:
APPLICANT: Seeliger, Hans Peter
APPLICANT: Renz, Manfred
TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
FILE REFERENCE: 8484-0043-999
CURRENT APPLICATION NUMBER: US/09/158,707
CURRENT FILING DATE: 1998-09-22
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 6328
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1) ... (5736)
US-09-158-707-1

Query Match
Best Local Similarity 83.2%; Score 15.8; DB 4; Length 6328;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19
Db 1453 TGGGGATCTCTGGAAGTGG 1435

RESULT 8
US-09-620-312D-325/c
Sequence 325, Application US/09620312D
Patent No. 6569662
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyun
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonhong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungting
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghaast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No.6569662el Nucleic Acids and
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D

CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: PL FL_genes Version 1.0
SEQ ID NO 325
LENGTH: 6475
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (219) ... (5957)
US-09-620-312D-325

Query Match
Best Local Similarity 83.2%; Score 15.8; DB 4; Length 6475;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19
Db 1671 TGGGGATCTCTGGAAGTGG 1653

RESULT 9
US-09-949-016-13910
Sequence 13910, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13910
LENGTH: 40897
TYPE: DNA
ORGANISM: Human
US-09-949-016-13910

Query Match
Best Local Similarity 83.2%; Score 15.8; DB 4; Length 40897;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19
Db 1452 TGGCGCTGCTGGAAGTGG 1470

RESULT 10
US-09-949-016-13911
Sequence 13911, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768

```

1 PRIOR FILING DATE: 2000-10-03
2 PRIOR APPLICATION NUMBER: 60/231,498
3 PRIOR FILING DATE: 2000-09-08
4 NUMBER OF SEQ ID NOS: 207012
5 SOFTWARE: FASTSEQ for Windows Version 4.0
6 SEQ ID NO 13911
7 LENGTH: 40897
8 TYPE: DNA
9 ORGANISM: Human
10 OS-09-949-016-13911

```

Query Match	83.2%	Score 15.8	DB 4	Length 40897
Best Local Similarity	89.5%	Pred. No. 1.8e+02		
Matches 17; Conservative	0	Mismatches 2	Indels 0	Gaps 0

DY 1 TGGGCTTCGTGAAGTGC 19
||| |||||||||
Db 1452 TGGCGTCCGTGAAGTGC 1470

```

RESULT 11
US-09-949-016-13912
; Sequence 13912, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13912
; LENGTH: 40857
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13912

```

Query Match	83.2%	Score 15.8;	DB 4;	Length 40897;
Best Local Similarity	89.5%;	Pred. No. 1.8e+02;		
Matches 17; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

QY 1 TGGGGTTCGCTGGAAGTCG 19
 |||||
 Db 1452 TGGCGGTCGCTGGAAGTCG 1470

```

RESULT 12
US-09-949-016-13913
; Sequence 13913, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 13913
; LENGTH: 40897
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13913

```

Query Match	83.2%	Score 15.8;	DB 4;	Length 40897;
Best Local Similarity	89.5%;	Pred. No. 1.8e+02;		
Matches 17; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

QY 1 TGGGGTTCGCTGAAGTGG 19
||| | | | | | | | | |
Db 1452 TGGCGTTCGCTGAAGTGG 1470

```

RESULT 13
US-09-949-016-14409
; Sequence 14409, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE. METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14409
; LENGTH: 40897
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14409

```

Query Match	83.2%	Score 15.8;	DB 4;	Length 40897;
Best Local Similarity	89.5%;	Pred. No. 1.8e+02;		
Matches 17;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0

QY 1 TGGGGTTCGCTGGAAGTGC 19
||| | ||||| |||||
Db 1452 TGGCGGTCGCTGGAAGTGC 1470

```

RESULT 14
US-09-949-016-14410
/
/ Sequence 14410, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949, 016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 14410
/ LENGTH: 40897
/ TYPE: DNA
/ ORGANISM: Human
/ US-09-949-016-14410

```

Query Match 83.2%; Score 15.8; DB 4; Length 40897;
 Best Local Similarity 89.5%; Pred. No. 1.8e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGGGTTGCTGGAAGTGG 19
 |||||
 DB 1452 TGGCGGTGCTGGAAGTGG 1470

RESULT 15

US-09-949-016-14411
 ; Sequence 14411, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 14411
 ; LENGTH: 40897
 ; TYPE: DNA
 ; ORGANISM: Human
 US-09-949-016-14411

Query Match 83.2%; Score 15.8; DB 4; Length 40897;
 Best Local Similarity 89.5%; Pred. No. 1.8e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGGGTTGCTGGAAGTGG 19
 |||||
 DB 1452 TGGCGGTGCTGGAAGTGG 1470

Search completed: November 6, 2005, 23:10:46
 Job time : 83.4787 secs

This page blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using BW model

Run on: November 6, 2005, 20:56:42 ; Search time 473.787 Seconds
(without alignments)
331.640 Million cell updates/sec

Title: US-10-777-131A-4

Sequence: 1 tggggctcgcgtgaagttg 19

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 22: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 23: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 24: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 25: /cgn2_6/ptodata/2/pubpna/US11_PUBCOMB.seq.*
- 26: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 27: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 28: /cgn2_6/ptodata/2/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	19	US-10-216-817-4	Sequence 4, Appli
2	19	100.0	19	US-10-777-131A-4	Sequence 4, Appli
3	19	100.0	1144	US-10-216-817-20	Sequence 20, Appli
4	19	100.0	1144	US-10-777-131A-20	Sequence 20, Appli
5	19	100.0	2406	US-10-216-817-32	Sequence 32, Appli

C 6	19	100.0	2406	22	US-10-777-131A-32	Sequence 32, Appli
C 7	19	100.0	2409	11	US-09-855-604-618	Sequence 618, App
C 8	19	100.0	2409	13	US-09-855-604-618	Sequence 618, App
C 9	19	100.0	2451	11	US-09-855-604-620	Sequence 620, App
C 10	19	100.0	2451	13	US-09-855-604-620	Sequence 620, App
C 11	19	100.0	2806	16	US-10-216-817-22	Sequence 22, Appli
C 12	19	100.0	2806	22	US-10-777-131A-22	Sequence 22, Appli
C 13	17.4	91.6	105077	21	US-10-417-375-15	Sequence 15, Appli
C 14	16.4	86.3	117328	22	US-10-461-862-127	Sequence 127, App
C 15	16	84.2	694	14	US-10-027-632-150246	Sequence 150246, Sequence 150246,
C 16	16	84.2	694	18	US-10-027-632-150246	Sequence 150246, Sequence 670362,
C 17	16	84.2	1343	13	US-09-925-065A-670362	Sequence 51483, A
C 18	15.8	83.2	201	22	US-10-741-600-51483	Sequence 51485, A
C 19	15.8	83.2	201	22	US-10-741-600-51485	Sequence 42574, A
C 20	15.8	83.2	255	19	US-10-424-559-42574	Sequence 22586, A
C 21	15.8	83.2	292	9	US-09-864-761-22586	Sequence 2257, App
C 22	15.8	83.2	304	10	US-09-764-891-2297	Sequence 40863, A
C 23	15.8	83.2	319	21	US-10-425-115-40863	Sequence 110341, A
C 24	15.8	83.2	403	21	US-10-425-115-110341	Sequence 96855, A
C 25	15.8	83.2	435	14	US-10-027-632-96855	Sequence 96855, A
C 26	15.8	83.2	435	18	US-10-027-632-96855	Sequence 5826, App
C 27	15.8	83.2	476	9	US-09-864-761-5826	Sequence 45152, A
C 28	15.8	83.2	584	13	US-09-925-065A-45152	Sequence 45153, A
C 29	15.8	83.2	584	13	US-09-925-065A-45153	Sequence 8112, App
C 30	15.8	83.2	628	20	US-10-767-701-8112	Sequence 490315, A
C 31	15.8	83.2	636	13	US-09-925-065A-490314	Sequence 490314, A
C 32	15.8	83.2	636	13	US-09-925-065A-490315	Sequence 20144, A
C 33	15.8	83.2	738	18	US-10-282-122A-20144	Sequence 682718, A
C 34	15.8	83.2	984	18	US-10-369-493-40669	Sequence 12370, A
C 35	15.8	83.2	1323	13	US-09-925-065A-682717	Sequence 2747, App
C 36	15.8	83.2	1323	13	US-09-925-065A-682718	Sequence 2747, App
C 37	15.8	83.2	2511	18	US-10-282-122A-12370	Sequence 12370, A
C 38	15.8	83.2	2725	18	US-10-320-797-22	Sequence 22, Appli
C 39	15.8	83.2	3240	18	US-10-104-047-170	Sequence 170, App
C 40	15.8	83.2	3268	24	US-10-450-763-23395	Sequence 23395, A
C 41	15.8	83.2	5739	9	US-09-960-253-142	Sequence 142, App
C 42	15.8	83.2	6417	9	US-09-962-436-288	Sequence 288, App
C 43	15.8	83.2	6417	17	US-10-096-534-28	Sequence 28, Appli
C 44	15.8	83.2	6417	22	US-10-843-641A-2747	Sequence 2747, App
C 45	15.8	83.2	6417	24	US-10-479-874A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-10-216-817-4
; Sequence 4, Application US/10216817
; Publication No. US20030129619A1
GENERAL INFORMATION:
APPLICANT: GICQUEL, BRIGITTE
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
FILE REFERENCE: 03495.0233-00000
CURRENT APPLICATION NUMBER: US/10/216,817
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: 60/311,824
PRIOR FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: 60/313,523
PRIOR FILING DATE: 2001-08-21
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 19
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Primer
US-10-216-817-4
Query Match 100.0%; Score 19; DB 16; Length 19;
Best Local Similarity 100.0%; Pred. No. 8.4;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19

Db 1 TGGGGTTCGCTGGAAGTGG 19

RESULT 2

US-10-777-131A-4

Sequence 4, Application US/10777131A

Publication No. US20050026216A1

GENERAL INFORMATION:

APPLICANT: GICOUEL, BRIGITTE

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG

TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS

TITLE OF INVENTION: IN GENES OF THE mult FAMILY

FILE REFERENCE: 02356.0090-00000

CURRENT FILING DATE: 2004-02-13

PRIOR APPLICATION NUMBER: US/10/777,131A

PRIOR FILING DATE: 2002-08-14

PRIOR APPLICATION NUMBER: 60/311,824

PRIOR FILING DATE: 2001-08-14

PRIOR APPLICATION NUMBER: 60/313,523

PRIOR FILING DATE: 2001-08-21

NUMBER OF SEQ ID NOS: 32

SOFTWARE: PatentIn Ver. 3.2

SEQ ID NO 4

LENGTH: 19

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

US-10-777-131A-4

Query Match 100.0%; Score 19; DB 22; Length 19;

Best Local Similarity 100.0%; Pred. No. 8.4;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19

Db 1 TGGGGTTCGCTGGAAGTGG 19

RESULT 3

US-10-216-817-20/c

Sequence 20, Application US/10216817

Publication No. US20030129619A1

GENERAL INFORMATION:

APPLICANT: GICOUEL, BRIGITTE

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG

TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS

TITLE OF INVENTION: IN GENES OF THE mult FAMILY

FILE REFERENCE: 03495.0233-00000

CURRENT FILING DATE: 2002-11-13

PRIOR APPLICATION NUMBER: US/10/216,817

PRIOR FILING DATE: 2001-08-14

PRIOR APPLICATION NUMBER: 60/313,523

PRIOR FILING DATE: 2001-08-21

NUMBER OF SEQ ID NOS: 32

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 20

LENGTH: 1144

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

US-10-216-817-20

Query Match 100.0%; Score 19; DB 16; Length 1144;

Best Local Similarity 100.0%; Pred. No. 7.5;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19

Db 1143 TGGGGTTCGCTGGAAGTGG 1125

RESULT 4

US-10-777-131A-20/c

Sequence 20, Application US/10777131A

Publication No. US20050026216A1

GENERAL INFORMATION:

APPLICANT: GICOUEL, BRIGITTE

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG

TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS

TITLE OF INVENTION: IN GENES OF THE mult FAMILY

FILE REFERENCE: 02356.0090-00000

CURRENT FILING DATE: 2004-02-13

PRIOR APPLICATION NUMBER: PCT/EP02/09679

PRIOR FILING DATE: 2002-08-14

PRIOR APPLICATION NUMBER: 60/311,824

PRIOR FILING DATE: 2001-08-14

PRIOR APPLICATION NUMBER: 60/313,523

PRIOR FILING DATE: 2001-08-21

NUMBER OF SEQ ID NOS: 32

SOFTWARE: PatentIn Ver. 3.2

SEQ ID NO 20

LENGTH: 1144

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

US-10-777-131A-20

Query Match 100.0%; Score 19; DB 22; Length 1144;

Best Local Similarity 100.0%; Pred. No. 7.5;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19

Db 1143 TGGGGTTCGCTGGAAGTGG 1125

RESULT 5

US-10-216-817-32/c

Sequence 32, Application US/10216817

Publication No. US20030129619A1

GENERAL INFORMATION:

APPLICANT: GICOUEL, BRIGITTE

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG

TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS

TITLE OF INVENTION: IN GENES OF THE mult FAMILY

FILE REFERENCE: 03495.0233-00000

CURRENT FILING DATE: 2002-11-13

PRIOR APPLICATION NUMBER: US/10/216,817

PRIOR FILING DATE: 2001-08-14

PRIOR APPLICATION NUMBER: 60/313,523

PRIOR FILING DATE: 2001-08-21

NUMBER OF SEQ ID NOS: 32

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 32

LENGTH: 2406

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

US-10-216-817-32

Query Match 100.0%; Score 19; DB 16; Length 2406;

Best Local Similarity 100.0%; Pred. No. 7.3;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGGGTTCGCTGGAAGTGG 19

Db 200 TGGGGTTCGCTGGAAGTGG 182

RESULT 6

US-10-777-131A-32/c
; Sequence 32, Application US/10777131A
; Publication No. US2005026216A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; FILE REFERENCE: 02356.0090-00000
; CURRENT APPLICATION NUMBER: US/10/777, 131A
; CURRENT FILING DATE: 2004-02-13
; PRIOR APPLICATION NUMBER: PCT/EP02/09679
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 32
; LENGTH: 2406
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-777-131A-32

Query Match 100.0%; Score 19; DB 22; Length 2406;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTGCTGGAAGTGG 19
|||||
DB 200 TGGGGTTGCTGGAAGTGG 182

RESULT 7
US-09-855-604-618/c
; Sequence 618, Application US/09855604
; Publication No. US20040214165A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: PORTNOI, DENIS
; APPLICANT: LIM, ENG-MONG
; APPLICANT: PELICIC, VLADIMIR
; APPLICANT: GUIGUENO, AGNES
; APPLICANT: GOGUET DE LA SALMONIERE, YVES
; TITLE OF INVENTION: POLYPEPTIDE NUCLEIC SEQUENCES EXPORTED FROM MYCOBACTERIA,
; TITLE OF INVENTION: VECTORS COMPRISING SAME AND USES FOR DIAGNOSING AND
; FILE REFERENCE: 03715.0062-01000
; CURRENT APPLICATION NUMBER: US/09/855,604
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/485,536
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: PCT/FR98/01813
; PRIOR FILING DATE: 1998-08-14
; PRIOR APPLICATION NUMBER: FR 97 10404
; PRIOR FILING DATE: 1997-08-14
; PRIOR APPLICATION NUMBER: FR 97 11325
; PRIOR FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 935
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 618
; LENGTH: 2409
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2406)
US-09-855-604-618

Query Match 100.0%; Score 19; DB 11; Length 2409;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTGCTGGAAGTGG 19
|||||
DB 200 TGGGGTTGCTGGAAGTGG 182

RESULT 8
US-09-855-604-618/c
; Sequence 618, Application US/09855604
; Publication No. US20050158714A3
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: PORTNOI, DENIS
; APPLICANT: LIM, ENG-MONG
; APPLICANT: PELICIC, VLADIMIR
; APPLICANT: GUIGUENO, AGNES
; APPLICANT: GOGUET DE LA SALMONIERE, YVES
; TITLE OF INVENTION: POLYPEPTIDE NUCLEIC SEQUENCES EXPORTED FROM MYCOBACTERIA,
; TITLE OF INVENTION: VECTORS COMPRISING SAME AND USES FOR DIAGNOSING AND
; FILE REFERENCE: 03715.0062-01000
; CURRENT APPLICATION NUMBER: US/09/855,604
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/485,536
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: PCT/FR98/01813
; PRIOR FILING DATE: 1998-08-14
; PRIOR APPLICATION NUMBER: FR 97 10404
; PRIOR FILING DATE: 1997-08-14
; PRIOR APPLICATION NUMBER: FR 97 11325
; PRIOR FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 935
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 618
; LENGTH: 2409
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2406)
US-09-855-604-618

Query Match 100.0%; Score 19; DB 13; Length 2409;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGGTTGCTGGAAGTGG 19
|||||
DB 200 TGGGGTTGCTGGAAGTGG 182

RESULT 9
US-09-855-604-620/c
; Sequence 620, Application US/09855604
; Publication No. US20040214165A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: PORTNOI, DENIS
; APPLICANT: LIM, ENG-MONG
; APPLICANT: PELICIC, VLADIMIR
; APPLICANT: GUIGUENO, AGNES
; APPLICANT: GOGUET DE LA SALMONIERE, YVES
; TITLE OF INVENTION: POLYPEPTIDE NUCLEIC SEQUENCES EXPORTED FROM MYCOBACTERIA,
; TITLE OF INVENTION: VECTORS COMPRISING SAME AND USES FOR DIAGNOSING AND
; FILE REFERENCE: 03715.0062-01000
; CURRENT APPLICATION NUMBER: US/09/855,604
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/485,536
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: PCT/FR98/01813
; PRIOR FILING DATE: 1998-08-14
; PRIOR APPLICATION NUMBER: FR 97 10404

```
/ PRIOR FILING DATE: 1997-08-14
/ PRIOR APPLICATION NUMBER: FR 97 11325
/ PRIOR FILING DATE: 1997-09-11
/ NUMBER OF SEQ ID NOS: 935
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO: 620
/ LENGTH: 2451
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (4)..(2448)
US-09-855-604-620
```

```
Query Match          100.0%; Score 19; DB 11; Length 2451;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 TGGGGTTCCGCTGGAAGTGG 19
         |||||
Db      242 TGGGGTTCCGCTGGAAGTGG 224
```

```
RESULT 10
US-09-855-604-620/c
/ Sequence 620, Application US/09855604
/ Publication No. US20050158714A9
/ GENERAL INFORMATION:
```

```
/ APPLICANT: GICQUEL, BRIGITTE
/ APPLICANT: PORTNOI, DENIS
/ APPLICANT: LIM, ENG-MONG
/ APPLICANT: PELICIC, VLADIMIR
/ APPLICANT: GUIGUENO, AGNES
/ APPLICANT: GOGUET DE LA SALMONIERE, YVES
/ TITLE OF INVENTION: POLYPEPTIDE NUCLEIC SEQUENCES EXPORTED FROM MYCOBACTERIA,
/ TITLE OF INVENTION: VECTORS COMPRISING SAME AND USES FOR DIAGNOSING AND
/ TITLE OF INVENTION: PREVENTING TUBERCULOSIS
/ FILE REFERENCE: 03715.0062-01000
/ CURRENT APPLICATION NUMBER: US/09/855,604
/ CURRENT FILING DATE: 2001-05-16
/ PRIOR APPLICATION NUMBER: 09/485,536
/ PRIOR FILING DATE: 2000-02-14
/ PRIOR APPLICATION NUMBER: PCT/FR98/01813
/ PRIOR FILING DATE: 1998-08-14
/ PRIOR APPLICATION NUMBER: FR 97 10404
/ PRIOR FILING DATE: 1997-08-14
/ PRIOR APPLICATION NUMBER: FR 97 11325
/ PRIOR FILING DATE: 1997-09-11
/ NUMBER OF SEQ ID NOS: 935
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO: 620
/ LENGTH: 2451
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (4)..(2448)
US-09-855-604-620
```

```
Query Match          100.0%; Score 19; DB 13; Length 2451;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 TGGGGTTCCGCTGGAAGTGG 19
         |||||
Db      242 TGGGGTTCCGCTGGAAGTGG 224
```

```
RESULT 11
US-10-216-817-22/c
/ Sequence 22, Application US/10216817
/ Publication No. US20030129619A1
/ GENERAL INFORMATION:
```

```
/ APPLICANT: GICQUEL, BRIGITTE
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
/ TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
/ TITLE OF INVENTION: IN GENES OF THE MULT FAMILY
/ FILE REFERENCE: 03495.0233-00000
/ CURRENT APPLICATION NUMBER: US/10/216,817
/ CURRENT FILING DATE: 2002-11-13
/ PRIOR APPLICATION NUMBER: 60/311,824
/ PRIOR FILING DATE: 2001-08-14
/ PRIOR APPLICATION NUMBER: 60/313,523
/ PRIOR FILING DATE: 2001-08-21
/ NUMBER OF SEQ ID NOS: 32
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO: 22
/ LENGTH: 2806
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
US-10-216-817-22
```

```
Query Match          100.0%; Score 19; DB 16; Length 2806;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 TGGGGTTCCGCTGGAAGTGG 19
         |||||
Db      400 TGGGGTTCCGCTGGAAGTGG 382
```

```
RESULT 12
US-10-777-131a-22/c
/ Sequence 22, Application US/10777131A
/ Publication No. US20050026216A1
/ GENERAL INFORMATION:
/ APPLICANT: GICQUEL, BRIGITTE
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
/ TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
/ TITLE OF INVENTION: IN GENES OF THE MULT FAMILY
/ FILE REFERENCE: 02356.0090-00000
/ CURRENT APPLICATION NUMBER: US/10/777,131A
/ CURRENT FILING DATE: 2004-02-13
/ PRIOR APPLICATION NUMBER: PCT/EP02/09679
/ PRIOR FILING DATE: 2002-08-14
/ PRIOR APPLICATION NUMBER: 60/311,824
/ PRIOR FILING DATE: 2001-08-14
/ PRIOR APPLICATION NUMBER: 60/313,523
/ PRIOR FILING DATE: 2001-08-21
/ NUMBER OF SEQ ID NOS: 32
/ SOFTWARE: PatentIn Ver. 3.2
/ SEQ ID NO: 22
/ LENGTH: 2806
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
US-10-777-131a-22
```

```
Query Match          100.0%; Score 19; DB 22; Length 2806;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 TGGGGTTCCGCTGGAAGTGG 19
         |||||
Db      400 TGGGGTTCCGCTGGAAGTGG 382
```

```
RESULT 13
US-10-417-375-15/c
/ Sequence 15, Application US/10417375
/ Publication No. US20040219528A1
/ GENERAL INFORMATION:
/ APPLICANT: David W. Morris
/ APPLICANT: Marc Malandro
/ TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
/ FILE REFERENCE: 529452001600
/ CURRENT APPLICATION NUMBER: US/10/417,375
```



```
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 105077
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(105077)
; OTHER INFORMATION: n = A,T,C or G
US-10-417-375-15
```

```
Query Match          91.6%; Score 17.4; DB 21; Length 105077;
Best Local Similarity 94.7%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 TGGGTTGCTGCTGGAAGTGG 19
DB      34989 TGGGTTGCTGCTGGAAGTGG 34971
```

```
RESULT 14
US-10-461-862-127/C
; Sequence 127, Application US/10461862
; Publication No. US20050904341
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001800
; CURRENT APPLICATION NUMBER: US/10/461,862
; CURRENT FILING DATE: 2003-06-13
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 127
; LENGTH: 117328
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(117328)
; OTHER INFORMATION: n = A,T,C or G
US-10-461-862-127
```

```
Query Match          86.3%; Score 16.4; DB 22; Length 117328;
Best Local Similarity 94.4%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      2 GGGGTTGCTGGAAGTGG 19
DB      14283 GGGGTTGCTGGAAGTGG 14266
```

```
RESULT 15
US-10-027-632-150246
; Sequence 150246, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
```

```
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150246
; LENGTH: 694
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-150246
```

```
Query Match          84.2%; Score 16; DB 14; Length 694;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      4 GGTTCGCTGGAAGTGG 19
DB      199 GGTTCGCTGGAAGTGG 214
```

Search completed: November 7, 2005, 05:21:14
Job time : 476.787 secs

This Page Blank (uspi01)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 02:38:52 ; Search time 1900.81 Seconds
(without alignments)
380.481 Million cell updates/sec

Title: US-10-777-131a-4

Perfect score: 19
Sequence: 1 tggggcttcgctggaagtg 19

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :
EST:
1: gb_est1:
2: gb_est2:
3: gb_hic:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gssi:
9: gb_gssi:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	17.4	91.6	495	5	BP055594 BP055594
C 2	17.4	91.6	508	1	AV778474 AV778474
C 3	17.4	91.6	512	5	BP028473 BP028473
C 4	17.4	91.6	519	1	AV779844 AV779844
C 5	17.4	91.6	523	1	AV779524 AV779524
C 6	17.4	91.6	526	4	BI419678 BI419678
C 7	17.4	91.6	528	1	AV779426 AV779426
C 8	17.4	91.6	531	1	AV778725 AV778725
C 9	17.4	91.6	532	1	BP043328 BP043328
C 10	17.4	91.6	536	1	AV777359 AV777359
C 11	17.4	91.6	552	1	AV777186 AV777186
C 12	17.4	91.6	557	5	BP055162 BP055162
C 13	17.4	91.6	565	5	BP055337 BP055337
C 14	17.4	91.6	565	5	BP056252 BP056252
C 15	17.4	91.6	583	5	BP034064 BP034064
C 16	17.4	91.6	604	1	AV776877 AV776877
C 17	17.4	91.6	607	1	AV777307 AV777307
C 18	17.4	91.6	638	8	AZ405854 AZ405854
C 19	17.4	91.6	785	5	BK626838 BK626838
C 20	17.4	91.6	915	4	BG178285 BG178285
C 21	17.4	91.6	960	9	CNS050735 CNS050735
C 22	16.4	86.3	354	6	BY550735 BY550735
C 23	16.4	86.3	413	5	BY310705 BY310705
C 24	16.4	86.3	490	5	BP057779 BP057779

C 25	16.4	86.3	510	5	BP052554 BP052554
C 26	16.4	86.3	517	5	BP053880 BP053880
C 27	16.4	86.3	526	5	BP065934 BP065934
C 28	16.4	86.3	706	1	AU089324 AU089324
C 29	16.4	86.3	742	2	BF143421 BF143421
C 30	16.4	86.3	922	9	CL507863 CL507863
C 31	16.4	86.3	1033	5	BQ925691 BQ925691
C 32	16.4	86.3	236	6	CB877779 CB877779
C 33	16.4	86.3	398	5	BQ759027 BQ759027
C 34	16.4	86.3	461	1	AV918034 AV918034
C 35	16.4	86.3	461	1	AV922614 AV922614
C 36	16.4	86.3	461	5	BQ660582 BQ660582
C 37	16.4	86.3	462	5	BQ664571 BQ664571
C 38	16.4	86.3	484	6	CA016770 CA016770
C 39	16.4	86.3	496	1	AV917430 AV917430
C 40	16.4	86.3	505	1	AV916508 AV916508
C 41	16.4	86.3	521	1	AV944743 AV944743
C 42	16.4	86.3	560	9	CL245534 CL245534
C 43	16.4	86.3	593	6	CB881023 CB881023
C 44	16.4	86.3	606	6	CB879649 CB879649
C 45	16.4	86.3	659	4	BJ227679 BJ227679

ALIGNMENTS

RESULT 1
BP055594/c 495 bp mRNA linear EST 25-AUG-2004
LOCUS BP055594
DEFINITION BP055594 Lotus corniculatus var. japonicus pods (less than 20 mm in length) Lotus corniculatus var. japonicus cDNA clone SPDL057g02_f
3', mRNA sequence.

ACCESSION BP055594
VERSION BP055594.1
KEYWORDS GI:45612345
SOURCE BP055594
ORGANISM Lotus corniculatus var. japonicus (Lotus japonicus)

REFERENCE Asanizu, E., Nakamura, Y., Sato, S., and Tabata, S.
TITLE Large-Scale Expressed Sequence Tag (EST) Analysis
JOURNAL Plant Mol. Biol. 54 (3), 405-414 (2004)
COMMENT Contact: Erika Asanizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asanizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source 1.495
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="SPDL057g02_f"
/issue_type="pods"
/dev_stage="pod (less than 20 mm in length)"
/clone_lib="Lotus corniculatus var. japonicus pods (less than 20 mm in length)"

ORIGIN

Query Match 91.6%; Score 17.4; DB 5; Length 495;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGGGTTCGCTGGAAGTG 19
|||||
Db 418 TGGGGTTCGCTGGAAGTG 400

RESULT 2
AV778474/c 508 bp mRNA linear EST 19-AUG-2004
LOCUS AV778474 Lotus japonicus Pods (20-30 mm in length) Lotus
DEFINITION corniculatus var. japonicus cDNA clone MPDL039C04_f 3', mRNA
sequence.
ACCESSION AV778474
VERSION AV778474.1 GI:45402131
KEYWORDS EST.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteeae;
Lotus.
1 (bases 1 to 508)
/db_xref="taxon:34305"
/clone="MPDL039C04_f"
/issue_type="Pods" (20-30 mm in length)"
/clone_lib="Lotus japonicus Pods (20-30 mm in length)"
/note="Vector: pBluescriptII SK-; Site_1: BcoRI; Site_2:
XhoI"
ORIGIN
Query Match 91.6%; Score 17.4; DB 1; Length 508;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGGGGTTGCTGGAAGTGG 19
|||||
DB 506 TGGGTTGGCTGGAAGTGG 488
RESULT 3
BP028473 512 bp mRNA linear EST 19-AUG-2004
LOCUS BP028473 Lotus corniculatus var. japonicus flower Lotus
DEFINITION corniculatus var. japonicus cDNA clone MP005a06_f 3', mRNA
sequence.
ACCESSION BP028473
VERSION BP028473.1 GI:45405633
KEYWORDS EST.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteeae;
Lotus.
1 (bases 1 to 512)
/db_xref="taxon:34305"
/clone="MPDL066B04_f"
/issue_type="Pods" (20-30 mm in length)"
/clone_lib="Lotus japonicus Pods (20-30 mm in length)"
/note="Vector: pBluescriptII SK-; Site_1: BcoRI; Site_2:
XhoI"
ORIGIN
Query Match 91.6%; Score 17.4; DB 1; Length 508;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGGGGTTGCTGGAAGTGG 19
|||||
DB 506 TGGGTTGGCTGGAAGTGG 488
JOURNAL Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1..508
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MPDL039C04_f"
/issue_type="Pods" (20-30 mm in length)"
/clone_lib="Lotus japonicus Pods (20-30 mm in length)"
/note="Vector: pBluescriptII SK-; Site_1: BcoRI; Site_2:
XhoI"
JOURNAL Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1..508
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MPDL039C04_f"
/issue_type="Pods" (20-30 mm in length)"
/clone_lib="Lotus japonicus Pods (20-30 mm in length)"
/note="Vector: pBluescriptII SK-; Site_1: BcoRI; Site_2:
XhoI"
JOURNAL Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1..508
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MPDL039C04_f"
/issue_type="Pods" (20-30 mm in length)"
/clone_lib="Lotus japonicus Pods (20-30 mm in length)"
/note="Vector: pBluescriptII SK-; Site_1: BcoRI; Site_2:
XhoI"

source 1..512
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MP005a06_f"
/issue_type="flower"
/clone_lib="Lotus corniculatus var. japonicus flower"
ORIGIN
Query Match 91.6%; Score 17.4; DB 5; Length 512;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGGGGTTGCTGGAAGTGG 19
|||||
DB 505 TGGGTTGGCTGGAAGTGG 487
RESULT 4
AV779844/c 519 bp mRNA linear EST 19-AUG-2004
LOCUS AV779844 Lotus japonicus Pods (20-30 mm in length) Lotus
DEFINITION corniculatus var. japonicus cDNA clone MPDL066B04_f 3', mRNA
sequence.
ACCESSION AV779844
VERSION AV779844.1 GI:45403495
KEYWORDS EST.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteeae;
Lotus.
1 (bases 1 to 519)
/db_xref="taxon:34305"
/clone="MPDL066B04_f"
/issue_type="Pods" (20-30 mm in length)"
/clone_lib="Lotus japonicus Pods (20-30 mm in length)"
/note="Vector: pBluescriptII SK-; Site_1: BcoRI; Site_2:
XhoI"
ORIGIN
Query Match 91.6%; Score 17.4; DB 1; Length 519;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGGGGTTGCTGGAAGTGG 19
|||||
DB 508 TGGGTTGGCTGGAAGTGG 490
JOURNAL Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1..519
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MPDL066B04_f"
/issue_type="Pods" (20-30 mm in length)"
/clone_lib="Lotus japonicus Pods (20-30 mm in length)"
/note="Vector: pBluescriptII SK-; Site_1: BcoRI; Site_2:
XhoI"
JOURNAL Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1..519
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MPDL066B04_f"
/issue_type="Pods" (20-30 mm in length)"
/clone_lib="Lotus japonicus Pods (20-30 mm in length)"
/note="Vector: pBluescriptII SK-; Site_1: BcoRI; Site_2:
XhoI"

```

VERSION      AV779524.1  GI:45403176
KEYWORDS
SOURCE       Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM     Lotus corniculatus var. japonicus
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Loteeae;
             Lotus.
REFERENCE    1 (bases 1 to 523)
             Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
             Characteristics of the Lotus japonicus Gene Repertoire Deduced from
             Large-Scale Expressed Sequence Tag (EST) Analysis
             Plant Mol. Biol. 54 (3), 405-414 (2004)
JOURNAL      Contact: Erika Asamizu
             The First Laboratory for Plant Gene Research
             Kazusa DNA Research Institute
             Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
             Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1..523
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MPDL059b01.f"
/tissue_type="Pods" (20-30 mm in length)
/clone_lib="Lotus japonicus Pods (20-30 mm in length)"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match      91.6%; Score 17.4; DB 1; Length 523;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGGGTTGCTGGAAGTGG 19
    |||||
Db 509 TGGGTTGCTGGAAGTGG 491

RESULT 6
BI419678      526 bp mRNA linear EST 15-AUG-2001
LOCUS         LJNSTR6d8r Lotus japonicus nodule library 5 and 7 week-old Lotus
DEFINITION    corniculatus var. japonicus cDNA 5', mRNA sequence.
ACCESSION     BI419678.1 GI:15190701
VERSION       BI419678.1
KEYWORDS
SOURCE        Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM      Lotus corniculatus var. japonicus
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Loteeae;
             Lotus.
REFERENCE    1 (bases 1 to 526)
             Colebatch,G., Freund,S., Trevaekis,B and Udvardi,M.
             Lotus japonicus root nodule ESTs: tools for functional genomics
             Unpublished (2000)
JOURNAL      Contact: Udvardi MK
             Molecular Plant Nutrition
             Max Planck Institute of Molecular Plant Physiology
             Am Muehlenberg 1, 14476 Golm, Germany
             Fax: 49 331 587 8250
             Email: udvardi@mpimp-golm.mpg.de
             Seq primer: T7
             High quality sequence sccp: 526.
FEATURES
source
1..526
/location/Qualifiers
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/cultivar="Gifu (B-129)"
/db_xref="taxon:34305"
/dev_stage="5 and 7 week-old plants"

/clone_lib="Lotus japonicus nodule library 5 and 7
week-old"
/notes="Organ: Nodule; Vector: pSPORT1; Site_1: SalI;
Site_2: NotI; The library was prepared using mRNA
extracted from nodules of 5 and 7 week-old Lotus plants.
Nodules were induced by, and contained Mesorhizobium
strain R7A."

ORIGIN
Query Match      91.6%; Score 17.4; DB 4; Length 526;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGGGTTGCTGGAAGTGG 19
    |||||
Db 75 TGGGTTGCTGGAAGTGG 93

RESULT 7
AV779426/c
LOCUS         AV779426 Lotus japonicus Pods (20-30 mm in length) Lotus
DEFINITION    corniculatus var. japonicus cDNA clone MPDL058b01.f 3', mRNA
sequence.
ACCESSION     AV779426
VERSION       AV779426
KEYWORDS
SOURCE        Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM      Lotus corniculatus var. japonicus
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Loteeae;
             Lotus.
REFERENCE    1 (bases 1 to 528)
             Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
             Characteristics of the Lotus japonicus Gene Repertoire Deduced from
             Large-Scale Expressed Sequence Tag (EST) Analysis
             Plant Mol. Biol. 54 (3), 405-414 (2004)
JOURNAL      Contact: Erika Asamizu
             The First Laboratory for Plant Gene Research
             Kazusa DNA Research Institute
             Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
             Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1..528
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MPDL058b01.f"
/tissue_type="Pods" (20-30 mm in length)
/clone_lib="Lotus japonicus Pods (20-30 mm in length)"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match      91.6%; Score 17.4; DB 1; Length 528;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGGGTTGCTGGAAGTGG 19
    |||||
Db 509 TGGGTTGCTGGAAGTGG 491

RESULT 8
AV778725/c
LOCUS         AV778725 Lotus japonicus Pods (20-30 mm in length) Lotus
DEFINITION    corniculatus var. japonicus cDNA clone MPDL044d02.f 3', mRNA
sequence.
ACCESSION     AV778725
VERSION       AV778725.1 GI:45402382
```

KEYWORDS EST.
SOURCE LOTUS corniculatus var. japonicus (Lotus japonicus)
ORGANISM LOTUS corniculatus var. japonicus

REFERENCE 1 (bases 1 to 531)
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE Characteristics of the Lotus japonicus Gene Repertoire Deduced from Large-Scale Expressed Sequence Tag (EST) Analysis
JOURNAL Plant Mol. Biol. 54 (3), 405-414 (2004)
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers

FEATURES
source
1..531
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MPDL04402.f"
/cisse_type="Pods (20-30 mm in length)"
/clone_1ib="Lotus japonicus Pods (20-30 mm in length)"
/note="Vector: pBluescriptII SK-, Site_1: EcoRI, Site_2: XhoI"

ORIGIN
Query Match 91.6%; Score 17.4; DB 1; Length 531;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGGGTTGCTGGAAGTGG 19
|||||
506 TGGGGTTGCTGGAAGTGG 488

Db 506 TGGGGTTGCTGGAAGTGG 488

RESULT 9
BP043328 532 bp mRNA linear EST 19-AUG-2004
LOCUS BP043328 Lotus corniculatus var. japonicus flower bud Lotus
DEFINITION corniculatus var. japonicus cDNA clone MPBL041c10_f 3', mRNA
sequence.
ACCESSION BP043328
VERSION BP043328.1 GI:45577097
KEYWORDS EST.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM LOTUS corniculatus var. japonicus

REFERENCE 1 (bases 1 to 532)
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE Characteristics of the Lotus japonicus Gene Repertoire Deduced from Large-Scale Expressed Sequence Tag (EST) Analysis
JOURNAL Plant Mol. Biol. 54 (3), 405-414 (2004)
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers

FEATURES
source
1..532
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MPBL041c10.f"
/cisse_type="flower bud"
/clone_1ib="Lotus corniculatus var. japonicus flower bud"

ORIGIN
Query Match 91.6%; Score 17.4; DB 5; Length 532;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGGGTTGCTGGAAGTGG 19
|||||
509 TGGGGTTGCTGGAAGTGG 491

Db 509 TGGGGTTGCTGGAAGTGG 491

RESULT 10
AV777359/c 536 bp mRNA linear EST 19-AUG-2004
LOCUS AV777359 Lotus japonicus Pods (20-30 mm in length) Lotus
DEFINITION corniculatus var. japonicus cDNA clone MPDL017e09_f 3', mRNA
sequence.
ACCESSION AV777359
VERSION AV777359.1 GI:45401024
KEYWORDS EST.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM LOTUS corniculatus var. japonicus

REFERENCE 1 (bases 1 to 536)
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE Characteristics of the Lotus japonicus Gene Repertoire Deduced from Large-Scale Expressed Sequence Tag (EST) Analysis
JOURNAL Plant Mol. Biol. 54 (3), 405-414 (2004)
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers

FEATURES
source
1..536
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MPDL017e09.f"
/cisse_type="Pods (20-30 mm in length)"
/clone_1ib="Lotus japonicus Pods (20-30 mm in length)"
/note="Vector: pBluescriptII SK-, Site_1: EcoRI, Site_2: XhoI"

ORIGIN
Query Match 91.6%; Score 17.4; DB 1; Length 536;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGGGTTGCTGGAAGTGG 19
|||||
508 TGGGGTTGCTGGAAGTGG 490

Db 508 TGGGGTTGCTGGAAGTGG 490

RESULT 11
AV777186 552 bp mRNA linear EST 19-AUG-2004
LOCUS AV777186 Lotus japonicus Pods (20-30 mm in length) Lotus
DEFINITION corniculatus var. japonicus cDNA clone MPDL013e05_f 3', mRNA
sequence.
ACCESSION AV777186
VERSION AV777186.1 GI:45400851
KEYWORDS EST.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM LOTUS corniculatus var. japonicus

REFERENCE 1 (bases 1 to 552)
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE Characteristics of the Lotus japonicus Gene Repertoire Deduced from Large-Scale Expressed Sequence Tag (EST) Analysis
JOURNAL Plant Mol. Biol. 54 (3), 405-414 (2004)
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers

FEATURES
source
1..552
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MPDL013e05.f"
/cisse_type="Pods (20-30 mm in length)"
/clone_1ib="Lotus japonicus Pods (20-30 mm in length)"
/note="Vector: pBluescriptII SK-, Site_1: EcoRI, Site_2: XhoI"

REFERENCE 1 (bases 1 to 552)
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE Characteristics of the Lotus japonicus Gene Repertoire Deduced from
JOURNAL Large-Scale Expressed Sequence Tag (EST) Analysis
COMMENT Plant Mol. Biol. 54 (3), 405-414 (2004)
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
FEATURES
source
1.552
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="SPDL050f03_f"
/issue_type="Pods (20-30 mm in length)"
/clone_1ib="Lotus japonicus Pods (20-30 mm in length)"
/note="Vector: pBluescriptII SK-, Site_1: EcoRI; Site_2:
XhoI"
ORIGIN
Query Match 91.6%; Score 17.4; DB 1; Length 552;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGGCGTTTCGCTGGAAGTGG 19
|||||
512 TGGCGTTTCGCTGGAAGTGG 494
RESULT 12
LOCUS BP055162/c 557 bp mRNA linear EST 25-AUG-2004
DEFINITION BP055162 Lotus corniculatus var. japonicus pods (less than 20 mm in
length) Lotus corniculatus var. japonicus cDNA clone SPDL050f03_f
3', mRNA sequence.
ACCESSION BP055162
VERSION BP055162.1 GI:45611913
KEYWORDS EST.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.
REFERENCE 1 (bases 1 to 557)
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE Characteristics of the Lotus japonicus Gene Repertoire Deduced from
JOURNAL Large-Scale Expressed Sequence Tag (EST) Analysis
COMMENT Plant Mol. Biol. 54 (3), 405-414 (2004)
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
FEATURES
source
1.557
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="SPDL050f03_f"
/issue_type="Pods (less than 20 mm in length)"
/dev_stage="pod (less than 20 mm in length)"
/clone_1ib="Lotus corniculatus var. japonicus pods (less
than 20 mm in length)"
ORIGIN
Query Match 91.6%; Score 17.4; DB 5; Length 557;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGTTTCGCTGGAAGTGG 19
|||||
478 TGGCGTTTCGCTGGAAGTGG 460
RESULT 13
LOCUS BP055337/c 565 bp mRNA linear EST 25-AUG-2004
DEFINITION BP055337 Lotus corniculatus var. japonicus pods (less than 20 mm in
length) Lotus corniculatus var. japonicus cDNA clone SPDL053c10_f
3', mRNA sequence.
ACCESSION BP055337
VERSION BP055337.1 GI:45612088
KEYWORDS EST.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.
REFERENCE 1 (bases 1 to 565)
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE Characteristics of the Lotus japonicus Gene Repertoire Deduced from
JOURNAL Large-Scale Expressed Sequence Tag (EST) Analysis
COMMENT Plant Mol. Biol. 54 (3), 405-414 (2004)
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
FEATURES
source
1.565
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="SPDL053c10_f"
/issue_type="Pods (less than 20 mm in length)"
/dev_stage="pod (less than 20 mm in length)"
/clone_1ib="Lotus corniculatus var. japonicus pods (less
than 20 mm in length)"
ORIGIN
Query Match 91.6%; Score 17.4; DB 5; Length 565;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGGCGTTTCGCTGGAAGTGG 19
|||||
456 TGGCGTTTCGCTGGAAGTGG 438
RESULT 14
LOCUS BP056252/c 565 bp mRNA linear EST 25-AUG-2004
DEFINITION BP056252 Lotus corniculatus var. japonicus pods (less than 20 mm in
length) Lotus corniculatus var. japonicus cDNA clone SPDL068h12_f
3', mRNA sequence.
ACCESSION BP056252
VERSION BP056252.1 GI:45613003
KEYWORDS EST.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.
REFERENCE 1 (bases 1 to 565)
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE Characteristics of the Lotus japonicus Gene Repertoire Deduced from
JOURNAL Large-Scale Expressed Sequence Tag (EST) Analysis
COMMENT Plant Mol. Biol. 54 (3), 405-414 (2004)

COMMENT

Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers

FEATURES

source

1..565
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="SPDL068h12.f"
/issue_type="pods"
/dev_stage="pod (less than 20 mm in length)"
/clone_1fb="Lotus corniculatus var. japonicus pods (less than 20 mm in length)"

ORIGIN

Query Match 91.6%; Score 17.4; DB 5; Length 565;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGGTTGCTGGAAGTGG 19
|||||
DB 486 TGGGTTGCTGGAAGTGG 468

RESULT 15 583 bp mRNA linear EST 19-AUG-2004
BP034064/c BP034064
LOCUS BP034064 Lotus corniculatus var. japonicus flower bud Lotus
DEFINITION corniculatus var. japonicus cDNA MFB002907_f3', mRNA
sequence.

ACCESSION BP034064 GI:45411224
VERSION BP034064.1
KEYWORDS EST.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.

REFERENCE 1 (bases 1 to 583)
Asamizu E, Nakamura Y, Sato S, and Tabata S.
Characteristics of the Lotus japonicus Gene Repertoire Deduced from
Large-Scale Expressed Sequence Tag (EST) Analysis
Plant Mol. Biol. 54 (3), 405-414 (2004)
JOURNAL Contact: Erika Asamizu
COMMENT The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers

FEATURES

source

1..583
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MFB002907.f"
/issue_type="flower bud"
/clone_1fb="Lotus corniculatus var. japonicus flower bud"

ORIGIN

Query Match 91.6%; Score 17.4; DB 5; Length 583;
Best Local Similarity 94.7%; Pred. No. 5.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGGTTGCTGGAAGTGG 19
|||||
DB 449 TGGGTTGCTGGAAGTGG 431

Search completed: November 7, 2005, 09:18:48

Job time : 1906.81 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 19:19:37 ; Search time 57.6915 Seconds
(without alignments)
482.163 Million cell updates/sec

Title: US-10-777-131a-5

Perfect score: 17

Sequence: 1 agccgcgtagtaacct 17

Scoring table:

IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_NA:*
1: /cgm2_6/prodata/1/ina/5A COMB.seq:*
2: /cgm2_6/prodata/1/ina/5B COMB.seq:*
3: /cgm2_6/prodata/1/ina/5A COMB.seq:*
4: /cgm2_6/prodata/1/ina/5B COMB.seq:*
5: /cgm2_6/prodata/1/ina/5A COMB.seq:*
6: /cgm2_6/prodata/1/ina/5B COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17	100.0	4403765	3 US-09-103-840A-2	Sequence 2, Appl1
C 2	17	100.0	4411529	3 US-09-103-840A-1	Sequence 1, Appl1
C 3	14	82.4	246	4 US-09-583-110-1035	Sequence 1035, Ap
C 4	14	82.4	246	4 US-09-583-110-1760	Sequence 1760, Ap
C 5	14	82.4	360	4 US-09-107-433-143	Sequence 143, Appl
C 6	14	82.4	1385	4 US-08-932-411A-17	Sequence 17, Appl
C 7	14	82.4	1385	4 US-09-585-645A-30	Sequence 30, Appl
C 8	14	82.4	1515	4 US-09-726-774-9	Sequence 9, Appl1
C 9	14	82.4	1522	4 US-09-548-606-1	Sequence 1, Appl1
C 10	14	82.4	6109	4 US-09-795-061-1	Sequence 20, Appl
C 11	14	82.4	21338	3 US-08-961-527-20	Sequence 121, Appl
C 12	13.8	81.2	793	2 US-08-967-101-121	Sequence 121, Appl
C 13	13.8	81.2	793	2 US-08-967-101-121	Sequence 121, Appl
C 14	13.8	81.2	793	3 US-09-124-698-121	Sequence 121, Appl
C 15	13.8	81.2	793	3 US-09-124-698-121	Sequence 121, Appl
C 16	13.8	81.2	793	3 US-08-496-841C-121	Sequence 121, Appl
C 17	13.8	81.2	793	3 US-09-124-523-121	Sequence 121, Appl
C 18	13.8	81.2	793	4 US-09-636-796A-121	Sequence 121, Appl
C 19	13.8	81.2	793	4 US-08-431-048F-121	Sequence 121, Appl
C 20	13.8	81.2	42931	4 US-08-311-731A-121	Sequence 121, Appl
C 21	13.8	81.2	107458	4 US-09-949-016-15687	Sequence 15687, A
C 22	13.8	81.2	203475	4 US-09-949-016-14517	Sequence 14517, A
C 23	13.8	81.2	203475	4 US-09-949-016-14517	Sequence 14517, A
C 24	13.8	81.2	203475	4 US-09-949-016-14517	Sequence 14517, A
C 25	13.8	81.2	203475	4 US-09-949-016-14519	Sequence 14519, A
C 26	13.8	81.2	203475	4 US-09-949-016-17226	Sequence 17226, A
C 27	13.8	81.2	203475	4 US-09-949-016-17227	Sequence 17227, A

C 28	13.8	81.2	203475	4 US-09-949-016-17228	Sequence 17228, A
C 29	13.8	81.2	203475	4 US-09-949-016-17229	Sequence 17229, A
C 30	13.4	78.8	969	4 US-09-252-991A-2630	Sequence 2630, Ap
C 31	13.4	78.8	1107	4 US-09-252-991A-2281	Sequence 2281, Ap
C 32	13.4	78.8	1596	4 US-09-489-039A-2610	Sequence 2610, Ap
C 33	13.4	78.8	2093	4 US-09-722-971-5	Sequence 5, Appl1
C 34	13.4	78.8	2943	1 US-08-042-747A-7	Sequence 7, Appl1
C 35	13.4	78.8	10059	4 US-09-949-016-15323	Sequence 15323, A
C 36	13	76.5	601	4 US-09-949-016-15323	Sequence 15323, A
C 37	13	76.5	87644	4 US-09-949-016-16041	Sequence 16041, A
C 38	12.8	75.3	51	4 US-09-443-199C-941	Sequence 941, Appl
C 39	12.8	75.3	51	4 US-09-443-199C-941	Sequence 941, Appl
C 40	12.8	75.3	52	3 US-09-291-874-27	Sequence 27, Appl
C 41	12.8	75.3	52	4 US-09-344-783C-44	Sequence 44, Appl
C 42	12.8	75.3	123	4 US-09-902-540-2904	Sequence 2904, Ap
C 43	12.8	75.3	378	4 US-09-252-991A-13252	Sequence 13252, A
C 44	12.8	75.3	471	4 US-09-902-540-6463	Sequence 6463, Ap
C 45	12.8	75.3	582	4 US-09-902-540-4043	Sequence 4043, Ap

ALIGNMENTS

```
RESULT 1
US-09-103-840A-2/c
; Sequence 2, Appl1action US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2
Query Match 100.0%; Score 17; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1478734 AGCCGCGTAGTAACCT 17
1 AGCCGCGTAGTAACCT 17
RESULT 2
US-09-103-840A-1/c
; Sequence 1, Appl1action US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
```

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 100.0%; Score 17; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGCTAGTAACCT 17
DB 1479210 AGCCGCTAGTAACCT 1479194

RESULT 3
US-09-583-110-1035/C
Sequence 1035, Application US/09583110
Patent No. 6699703
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al.
TITLE OF INVENTION: Nucleic Acid Sequences Relating to Streptococcus
TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
FILE REFERENCE: PAT00-07A
CURRENT APPLICATION NUMBER: US/09/583,110
CURRENT FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/107,433
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/085,131
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: US 60/051,553
PRIOR FILING DATE: 1997-07-02
NUMBER OF SEQ ID NOS: 5322
SEQ ID NO 1035
LENGTH: 246
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-583-110-1035

Query Match 82.4%; Score 14; DB 4; Length 246;
Best Local Similarity 100.0%; Pred. No. 2,6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CGCGTAGTAACCT 17
DB 131 CGCGTAGTAACCT 118

RESULT 4
US-09-583-110-1760/C
Sequence 1760, Application US/09583110
Patent No. 6699703
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al.
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
FILE REFERENCE: PAT00-07A
CURRENT APPLICATION NUMBER: US/09/583,110
CURRENT FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/107,433
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/085,131
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: US 60/051,553
PRIOR FILING DATE: 1997-07-02
NUMBER OF SEQ ID NOS: 5322
SEQ ID NO 1760
LENGTH: 246
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-583-110-1760

Query Match 82.4%; Score 14; DB 4; Length 246;
Best Local Similarity 100.0%; Pred. No. 2,6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CGCGTAGTAACCT 17
DB 131 CGCGTAGTAACCT 118

RESULT 5
US-09-107-433-143/C
Sequence 143, Application US/09107433
Patent No. 6800744
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGN
THERAPEUTICS
NUMBER OF SEQUENCES: 5206
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 143:
SEQUENCE CHARACTERISTICS:
LENGTH: 360 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULAR TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (8) LOCATION 1...360
SEQUENCE DESCRIPTION: SEQ ID NO: 143:
US-09-107-433-143

Query Match 82.4%; Score 14; DB 4; Length 360;
Best Local Similarity 100.0%; Pred. No. 2,6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CGCGTAGTAACCT 17
DB 49 CGCGTAGTAACCT 36

```
RESULT 6
US-08-932-411A-17
; Sequence 17, Application US/08932411A
; Patent No. 6566496
; GENERAL INFORMATION:
; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qifu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/932,411A
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/772,009
; FILING DATE: 19-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/722,570
; FILING DATE: 19-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1385 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 382..1170
; US-08-932-411A-17

Query Match      82.4%; Score 14; DB 4; Length 1385;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCGCGTAGGTAA 14
      |||
Db      104 AGCCGCGTAGGTAA 117

RESULT 7
US-09-585-645A-30
; Sequence 30, Application US/09585645A
; Patent No. 6838444
; GENERAL INFORMATION:
; APPLICANT: Zoghbi, Huda
; APPLICANT: Bellan, Hugo
; APPLICANT: Birmingham, Neesam
; APPLICANT: Hassan, Beesam
; APPLICANT: Ben-Arie, Nissim
; TITLE OF INVENTION: Compositions and Methods for Therapeutic Use of Atonal-associated
; TITLE OF INVENTION: for Deafness, Osteoarthritis, and Abnormal Cell Proliferation
; FILE REFERENCE: P01899052
```

```
; CURRENT APPLICATION NUMBER: US/09/585,645A
; CURRENT FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: US 60/137,060
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: US 60/176,993
; PRIOR FILING DATE: 2000-01-19
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: Patent version 3.1
; SEQ ID NO 30
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: MOUSE
US-09-585-645A-30

Query Match      82.4%; Score 14; DB 4; Length 1385;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCGCGTAGGTAA 14
      |||
Db      104 AGCCGCGTAGGTAA 117

RESULT 8
US-09-726-774-9
; Sequence 9, Application US/09726774
; Patent No. 6677153
; GENERAL INFORMATION:
; APPLICANT: Iversen, Patrick L.
; TITLE OF INVENTION: Antisense Antibacterial Method and
; FILE REFERENCE: 0450-0032.30
; CURRENT APPLICATION NUMBER: US/09/726,774
; CURRENT FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: US 60/168,150
; PRIOR FILING DATE: 1999-11-29
; NUMBER OF SEQ ID NOS: 139
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1515
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-726-774-9

Query Match      82.4%; Score 14; DB 4; Length 1515;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 CGCGTAGTAACT 17
      |||
Db      107 CGCGTAGTAACT 120

RESULT 9
US-09-548-606-1
; Sequence 1, Application US/09548606
; Patent No. 6579711
; GENERAL INFORMATION:
; APPLICANT: GAIER, Walter et al.
; TITLE OF INVENTION: NOVEL LACTIC ACID BACTERIA SPECIES
; FILE REFERENCE: 8265-322
; CURRENT APPLICATION NUMBER: US/09/548,606
; CURRENT FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: PCT/EP98/06636
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 97203245.2
; PRIOR FILING DATE: 1997-10-17
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1522
; TYPE: DNA
; ORGANISM: Bacteria
```

FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(1522)
OTHER INFORMATION: n = A,T,C or G
US-09-548-606-1

Query Match 82.4%; Score 14; DB 4; Length 1522;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CGCGTAGGTAACT 17
|||||
DB 118 CGCGTAGGTAACT 131

RESULT 10
US-09-795-061-1/C
Sequence 1, Application US/09795061
Patent No. 6759528

GENERAL INFORMATION:

APPLICANT: Greenspan, Daniel S

APPLICANT: Imamura, Yasutada

TITLE OF INVENTION: Pro-Alpha 3 (V) Collagen Genes

FILE REFERENCE: 960296.96781

CURRENT APPLICATION NUMBER: US/09/795,061

CURRENT FILING DATE: 2001-02-26

NUMBER OF SEQ ID NOS: 24

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 1

LENGTH: 6109

TYPE: DNA

ORGANISM: Mus musculus

FEATURE:

NAME/KEY: CDS

LOCATION: (82)..(5298)

US-09-795-061-1

Query Match 82.4%; Score 14; DB 4; Length 6109;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCGCGTAGGTAA 14
|||||

DB 6022 AGCGCGTAGGTAA 6009

RESULT 11
US-08-961-527-20
Sequence 20, Application US/08961527
Patent No. 6420135

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:

ADDRESSER: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8504

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 2138 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-20

Query Match 82.4%; Score 14; DB 3; Length 2138;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CGCGTAGGTAACT 17
|||||

DB 20823 CGCGTAGGTAACT 20836

RESULT 12
US-08-967-101-121
Sequence 121, Application US/08967101
Patent No. 5840540

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H

APPLICANT: ROMMENS, JOHANNA M

APPLICANT: FRASER, PAUL E

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TITLE OF INVENTION: TO ALZHEIMER'S DISEASE

NUMBER OF SEQUENCES: 183

CORRESPONDENCE ADDRESS:

ADDRESSER: TESTA, HURWITZ & THIBEAULT

STREET: High Street Tower - 125 High Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/967,101

FILING DATE: 10-NOV-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/592,541

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pletcher, Edmund R.

TELEPHONE: (617) 248-7100

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 121:

SEQUENCE CHARACTERISTICS:

LENGTH: 793 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-967-101-121

Query Match 81.2%; Score 13.8; DB 2; Length 793;
Best Local Similarity 88.2%; Pred. No. 3.4e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGCGCGTAGGTAACT 17

Db 7 AGCCGCGTATTACCT 23

RESULT 13

US-08-592-541-121
Sequence 121, Application US/08592541
Patent No. 5986054

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M

APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,541

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 121:

SEQUENCE CHARACTERISTICS:

LENGTH: 793 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-592-541-121

Query Match 81.2%; Score 13.8; DB 2; Length 793;

Best Local Similarity 88.2%; Pred. No. 3.4e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGCCGCGTATTACCT 17
Db 7 AGCCGCGTATTACCT 23

RESULT 14

US-09-124-698-121

Sequence 121, Application US/09124698

Patent No. 6117978

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H

APPLICANT: ROMMENS, JOHANNA M

APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT

STREET: High Street Tower - 125 High Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/124,698

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/592,541

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 121:

SEQUENCE CHARACTERISTICS:

LENGTH: 793 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-124-698-121

Query Match 81.2%; Score 13.8; DB 3; Length 793;

Best Local Similarity 88.2%; Pred. No. 3.4e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGCCGCGTATTACCT 17
Db 7 AGCCGCGTATTACCT 23

RESULT 15

US-09-127-480-121

Sequence 121, Application US/09127480

Patent No. 6194153

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H

APPLICANT: ROMMENS, JOHANNA M

APPLICANT: FRASER, PAUL E

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT

STREET: High Street Tower - 125 High Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/127,480

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/592,541

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 121:

SEQUENCE CHARACTERISTICS:

LENGTH: 793 base pairs

TYPE: nucleic acid

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-127-480-121

```

```

Query Match      81.2%; Score 13.8; DB 3; Length 793;
Best Local Similarity 88.2%; Pred. No. 3.4e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      1 AGCCGCGTAGTACCT 17
        |||||
        |||||
DB      7 AGCCGCGTATTACCT 23

```

Search completed: November 6, 2005, 23:11:06
Job time : 77.6915 secs

REFERENCE Bacteria; Actinobacteria; Actinomycetales;
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium tuberculosis complex.
1
AUTHORS Gicquel, B.
TITLE Compositions and methods for detecting multidrug resistant strains of M. tuberculosis having mutations in genes of the mtrT family
JOURNAL Patent: WO 03016562-A 17 27-FEB-2003;
INSTITUT PASTEUR (FR)
FEATURES Location/Qualifiers
source 1. 2488
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"
ORIGIN
Query Match 100.0%; Score 17; DB 6; Length 2488;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGCCGCGTAGTAACCT 17
|||||
Db 407 AGCCGCGTAGTAACCT 423
RESULT 3
MSGY151 37036 bp DNA linear BCT 10-DEC-1996
LOCUS Mycobacterium tuberculosis sequence from clone y151.
DEFINITION AD000018
ACCESSION AD000018
VERSION AD000018.1 GI:11717734
KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinomycetales;
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium tuberculosis complex.
1 (bases 1 to 37036)
REFERENCE
AUTHORS Du, L.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1996) L. Du, Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA, USA, 02154 du@critic.com
COMMENT GSDB:S:1004706.
FEATURES Location/Qualifiers
source 1. 37036
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"
/clone="y151"
ORIGIN
Query Match 100.0%; Score 17; DB 1; Length 37036;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGCCGCGTAGTAACCT 17
|||||
Db 36091 AGCCGCGTAGTAACCT 36107
RESULT 4
AE000516_14/c
WPCOMMENT
Sequence split into 44 fragments LOCUS AE000516 Accession AE000516
Fragment Name Begin End
AE000516_00 1 110000
AE000516_01 100001 210000
AE000516_02 200001 310000
AE000516_03 300001 410000
AE000516_04 400001 510000
AE000516_05 500001 610000
AE000516_06 600001 710000
AE000516_07 700001 810000

AE000516_08 800001 910000
AE000516_09 900001 1010000
AE000516_10 100001 1110000
AE000516_11 110001 1210000
AE000516_12 120001 1310000
AE000516_13 130001 1410000
AE000516_14 140001 1510000
AE000516_15 150001 1610000
AE000516_16 160001 1710000
AE000516_17 170001 1810000
AE000516_18 180001 1910000
AE000516_19 190001 2010000
AE000516_20 200001 2110000
AE000516_21 210001 2210000
AE000516_22 220001 2310000
AE000516_23 230001 2410000
AE000516_24 240001 2510000
AE000516_25 250001 2610000
AE000516_26 260001 2710000
AE000516_27 270001 2810000
AE000516_28 280001 2910000
AE000516_29 290001 3010000
AE000516_30 300001 3110000
AE000516_31 310001 3210000
AE000516_32 320001 3310000
AE000516_33 330001 3410000
AE000516_34 340001 3510000
AE000516_35 350001 3610000
AE000516_36 360001 3710000
AE000516_37 370001 3810000
AE000516_38 380001 3910000
AE000516_39 390001 4010000
AE000516_40 400001 4110000
AE000516_41 410001 4210000
AE000516_42 420001 4310000
AE000516_43 430001 4400000
Continuation (15 of 44) of AE000516 from base 1400001 (AE000516 Mycobacterium tuberculosis
Query Match 100.0%; Score 17; DB 1; Length 110000;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGCCGCGTAGTAACCT 17
|||||
Db 78753 AGCCGCGTAGTAACCT 78737
RESULT 5
BX248338 299450 bp DNA linear BCT 06-MAY-2004
LOCUS Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment 5/14.
DEFINITION BX248338 BX248333
ACCESSION BX248338.1 GI:31617962
VERSION
KEYWORDS complete genome.
SOURCE Mycobacterium bovis AF2122/97
ORGANISM Mycobacterium bovis AF2122/97
Bacteria; Actinobacteria; Actinomycetales;
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium tuberculosis complex.
1
REFERENCE
AUTHORS Garnier, T., Eiglmeyer, K., Camus, J.-C., Medina, N., Mansoor, H., Pryor, M., Dutkov, S., Gordin, S., Lacroix, C., Monsemp, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P.R., Parkhill, J., Barrell, B.G., Cole, S.T., Gordon, S.V. and Hewinson, G.
TITLE The complete genome sequence of Mycobacterium bovis
JOURNAL Unpublished
AUTHORS Garnier, T.
REFERENCE Direct Submission
TITLE Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
JOURNAL Bacterieme Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex

PROTEIN MPE110"
/protein_id="CAD94076.1"
/db_xref="GI:31617965"
/db_xref="GOA:O50439"
/db_xref="UniProt/Swiss-Prot:O50439"
/translation="WGCNVALALVLPVAVPSLMAQORHVPVVPADAPSVAVROM
AEAFHSGSENLVLLTDEKGLADENVLTVDRLENDKDVLMQDPTTPELR
EVLGSKDGKAWLPIGLAGDLTPKSYHAYTVVERIVKRTAGTLLTNAVGPATVA
DLTDAGARBARIBELAVMLVILMVITRNPTWMLPLVTIGASLMTAQAQVAVGLQ
VGLAVSNOAIVLSAMTAGQDYAVFLISYHEXVRGEPREAVORAMNSGVKI
AASAACTGTTCGMRPAKIGVSTGPALAIIGIAPLAVALLPALIVLAPRGMVA
RGERMATMRKARGRIRPRAYIGASLIGVALASCAHAFVVDKCOLPSPDPS
RNGVAMEHHSFVNQITPEYLIHSHMDIKTRIGLADLEQLAKORSQIPGVAMGVT
RNGETLEQARATYQAGQGNRLGASRMIDERTDNLRLASGANLADNEDVARGQV
SRAVGSRLVDAVYIQNPGKNTFENIDNARLVNSIHALGALQVDFGILANSF
DLDSVVALDTPSCVDSNPGCNARVCPHKLQTAEDNGTLDKRVGLAQOLSTSPQ
TVSAVNDIGRSINSVYRSLKSLGLNDPAAARALISONGANDILASAGROVADGVQ
LVDQTNMGTGALQASAFIMAGNDASQSMAGFVPPQVLSSEFKKVAQAFISPDG
HTVRYPITQDLPFSTAAADQVNTIITDKAGOPTSLDASISMSGYVMLRDIDY
YERDMRLIVAVTVVVLILMLALRAIVAPLVGVSVIYSMSAIGLVAVVQVEVGLQ
ELHMSVGLAFVVLVAVGADYMLLASRLDSALGRSVIRVACTGAVTTAGLI
FASMSGLEFSIGTVGGFTIGVILIDTFVVRITVPAMATLLGRASWMPGHPWQ
RPAEFGCMASMSARTKTVPOAVDGSKR"
complement(11182..12261)
/locus_tag="Mb1216c"
complement(11182..12261)
/locus_tag="Mb1216c"
/note="Mb1216c", len: 359 aa. Equivalent to Rv1184c,
len: 359 aa, from Mycobacterium tuberculosis strain H37Rv,
(100% identity in 359 aa overlap). Possible exported

Query Match 100.0%; Score 17; DB 1; Length 299450;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACCT 17
|||||
Db 162647 AGCCGCGTAGTAACCT 162631

RESULT 6
BX842576/c 348264 bp DNA linear BCF 10-JUN-2004
LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 5/13.
DEFINITION BX842576 AL021006 AL021184 AL123456 Z73419 Z73902 Z75555
ACCESSION Z77137 Z77164 Z77826 Z79701 Z80108 Z81011 Z83862 Z95844
VERSION BX842576.1 GI:38490250
KEYWORDS complete genome.
SOURCE Mycobacterium tuberculosis H37Rv
ORGANISM Mycobacterium tuberculosis; Actinobacteridae; Actinomycetales;
Bacteria; Actinobacteria; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
Teale, F., Badcock, K., Braham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squares, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.
TITLE Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
JOURNAL Nature 393 (6685), 537-544 (1998)
MEDLINE 98295987
PUBMED 9634230
REFERENCE
AUTHORS Camus, J.C., Pryor, M.J., Medigue, C. and Cole, S.T.
TITLE Re-annotation of the genome sequence of Mycobacterium tuberculosis
H37Rv
JOURNAL Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)

MEDLINE 22255591
PUBMED 12368430
REFERENCES 3 (bases 1 to 348264)
AUTHORS Parkhill, J.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unité de Génétique
Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On or before Nov 21, 2003 This sequence version replaced
gi:3242277, gi:3261543, gi:3261541, gi:3261573, gi:3261576,
gi:3261608, gi:3261598, gi:3261615, gi:3261623, gi:3261635,
gi:3256012, gi:3242274, gi:3261684, gi:3250713.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the world wide web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/).

FEATURES
source
1.348264
/organism="Mycobacterium tuberculosis H37Rv"
/mol_type="genomic DNA"
/strain="H37Rv"
/db_xref="taxon:83332"
170..1159
/gene="mdh"
/locus_tag="Rv1240"
170..1159
/gene="mdh"
/locus_tag="Rv1240"
/function="INVOLVED IN THE CONVERSION OF MALATE TO
OXALACETATE [CATALYTIC ACTIVITY: (S)-malate + NAD+ =
oxaloacetate + NADH]."
/note="Rv1240, (MTV006.12), len: 329 aa. Probable mdh,
Malate dehydrogenase (EC 1.1.1.37). Most similar to
P50917|MDH MYCIE MALATE DEHYDROGENASE from Mycobacterium
leprae (325 aa), FASTA scores: opt: 1887, E(): 0, (87.1%
identity in 329 aa overlap). Contains P500068 Malate
dehydrogenase active site signature. BELONGS TO THE LDH
FAMILY. MDH SUBFAMILY. Tbpase score is 0.867."
/codon_start=1
/evidence=experimental
/transl_table=11
/product="PROBABLE MALATE DEHYDROGENASE MDH"
/protein_id="CAA15896.1"
/db_xref="GI:2695826"
/translation="MSASPLKVAVTGAQGIYSLIFRLASGSLGPDRFIEURLLEI
EPALQALEGVMEIDDCAPFLSGVEIGSPDKTIFGVSLALLVGAAPRQAGERSDL
LEANGAIFTAQKALNVAADVIVGVGSPATNALIAMTNAPDIPIRERFASLRLD
HNRAISQALKAAGAAVTDIKKMTWGNHSAIOYDPEFAEVAKNAAEVNDQAWIED
EPIPTAKAGAAIITDARGASASASATIDARDWMLIGRPADWMSAVVSGSYCV
PEGLISSFPVTTGKMWTVISGLEIDFBRGRIDKSTAEADRSAVTELGTL"
635..673
/gene="mdh"
/locus_tag="Rv1240"
/note="P500068 Malate dehydrogenase active site signature"
1235..1495
/locus_tag="Rv1241"
1235..1495
/locus_tag="Rv1241"
/function="UNKNOWN"
/note="Rv1241, (MTV006.13), len: 86 aa. Conserved
hypothetical protein, member of family of 16 hypothetical
Mycobacterium tuberculosis proteins including: PROTEIN C12
Rv2871|Q10799|Y871 MYCTU HYPOTHETICAL 13.2 kDa
(124 aa), FASTA scores: opt: 172, E(): 9.5e-06, (37.2%
identity in 86 aa overlap); Rv2132, Rv3321c, etc. Tbpase
score is 0.875."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAA15897.1"

```
misc_feature
      3850..3882
        /gene=".1pgz"
        /locus_tag="RV1244"
        /note="P800013 Prokaryotic membrane lipoprotein lipid attachment site"
        complement(475..558)
        /locus_tag="RV1245c"
        complement(475..558)
        /locus_tag="RV1245c"
        /EC_number="1.-.-.-"
        /function="UNKNOWN; SUPPOSED INVOLVED IN CELLULAR METABOLISM."
        /note="RV1245C, (MTY006.17C), len: 276 aa. Probable short-chain dehydrogenase/reductase (EC 1.-.-.-), equivalent to NP_301801.1|NC_002677 short chain alcohol dehydrogenase from Mycobacterium leprae (277 aa). Also highly similar to various dehydrogenases and oxidoreductases e.g. NP_250228.1|NC_002516 probable short-chain dehydrogenase from Pseudomonas aeruginosa (295 aa); NP_421969.1|NC_002696 short chain dehydrogenase family protein from Caulobacter crescentus (278 aa); etc. Also highly similar to others from Mycobacterium
```


clone end:Sp6
site:EcoRI
end_sequence:BH323125"

ORIGIN

Query Match 90.6%; Score 15.4; DB 2; Length 221666;
Best Local Similarity 94.1%; Pred. No. 5.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AGCCGGTAGTACCT 17
|||||
Db 185623 AGCCGGTAGTACCT 185639

RESULT 10

AC122442/c 180516 bp DNA linear ROD 27-NOV-2003
LOCUS Mus musculus BAC clone RP24-230D14 from chromosome 19, complete
DEFINITION sequence.

AC122442 AC122442 GI:34556312
VERSION AC122442.4
KEYWORDS HTG.

SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 180516)
Hodges,J. and Bielicki,L.

2 (bases 1 to 180516)
The sequence of Mus musculus BAC clone RP24-230D14

3 (bases 1 to 180516)
Unpublished (2001)

4 (bases 1 to 180516)
Sequencing of Mus musculus

5 (bases 1 to 180516)
Unpublished (2001)

6 (bases 1 to 180516)
McPherson,J.D. and Waterston,R.H.

7 (bases 1 to 180516)
Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

8 (bases 1 to 180516)
McPherson,J.D. and Waterston,R.H.

9 (bases 1 to 180516)
Submitted (01-APR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

10 (bases 1 to 180516)
Wilson,R.K.

11 (bases 1 to 180516)
Submitted (10-SEP-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

12 (bases 1 to 180516)
Wilson,R.

13 (bases 1 to 180516)
Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

14 (bases 1 to 180516)
On Sep 10, 2003 this sequence version replaced gi:29423952.

15 (bases 1 to 180516)
Center: Washington University Genome Sequencing Center

16 (bases 1 to 180516)
Center code: WUSGC

17 (bases 1 to 180516)
Web site: http://genome.wustl.edu

18 (bases 1 to 180516)
Contact: submissions@wustl.wustl.edu

19 (bases 1 to 180516)
Summary Statistics

20 (bases 1 to 180516)
Center project name: W_B80230D14

21 (bases 1 to 180516)
Center project name: W_B80230D14

22 (bases 1 to 180516)
Center project name: W_B80230D14

23 (bases 1 to 180516)
Center project name: W_B80230D14

24 (bases 1 to 180516)
Center project name: W_B80230D14

25 (bases 1 to 180516)
Center project name: W_B80230D14

26 (bases 1 to 180516)
Center project name: W_B80230D14

as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:

The RPI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES

source

1..180516

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="19"

/map="19"

/clone="RP24-230D14"

/clone_1fb="RP24-230D14"

177..287

/rpt_family="A1u"

732..908

/rpt_family="B2"

961..1100

/rpt_family="B2"

1395..1600

/rpt_family="B2"

1703..1891

/rpt_family="B2"

2014..2075

/rpt_family="A1u"

2175..2378

/rpt_family="B2"

2394..2538

/rpt_family="A1u"

2946..3048

/rpt_family="B4"

3428..3622

/rpt_family="B2"

3751..3805

/rpt_family="BRY1"

3853..3991

/rpt_family="L1"

4004..4076

/rpt_family="ID"

4191..4300

/rpt_family="A1u"

4403..4596

/rpt_family="B2"

4623..4782

/rpt_family="B4"

4784..4821

/rpt_family="L1"

4892..5083

/rpt_family="B2"

5097..5280

/rpt_family="B2"

5564..5711

/rpt_family="L1"

5827..5919

/rpt_family="L1"

5969..6099

/rpt_family="A1u"

7028..7162

/rpt_family="B2"

/rpt_family="B2"

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

```

repeat_region      7344..7456
                    /rpt_family="Alu"
repeat_region      8076..8212
                    /rpt_family="Alu"
repeat_region      8810..8914
                    /rpt_family="Alu"
repeat_region      9356..9875
                    /rpt_family="RMR19B"
repeat_region      9869..10093
                    /rpt_family="RMR19B"
repeat_region      10094..10266
                    /rpt_family="B2"
repeat_region      10267..10403
                    /rpt_family="RMR19B"
repeat_region      10936..11126
                    /rpt_family="B2"
repeat_region      11379..11465
                    /rpt_family="Alu"
repeat_region      14030..14042
                    /rpt_family="L1"
repeat_region      14043..14239
                    /rpt_family="B2"
repeat_region      14240..15016
                    /rpt_family="L1"
repeat_region      15740..15807
                    /rpt_family="ID"
repeat_region      15884..15982
                    /rpt_family="B4"
repeat_region      16454..16513
                    /rpt_family="B4"
repeat_region      16514..16691
                    /rpt_family="B2"
repeat_region      16724..16932
                    /rpt_family="B4"
repeat_region      17027..17241
                    /rpt_family="B2"
repeat_region      17507..17582
                    /rpt_family="ID"
repeat_region      17880..17975
                    /rpt_family="Alu"
repeat_region      17986..18056
                    /rpt_family="ID"
repeat_region      18118..18181
                    /rpt_family="Alu"
repeat_region      18135..18147
                    /note="Sequence derived from one plasmid subclone."
unsure
repeat_region      18295..18330
                    /rpt_family="Alu"
repeat_region      18306..18393
                    /rpt_family="B4"
repeat_region      19340..19363
                    /rpt_family="B2"
repeat_region      19364..19512
                    /rpt_family="Alu"
repeat_region      19513..19680
                    /rpt_family="B2"
repeat_region      19905..20125
                    /rpt_family="B4"
repeat_region      20602..20745
                    /rpt_family="B2"
repeat_region      20915..21176
                    /rpt_family="B4"
repeat_region      21394..21441
                    /rpt_family="Alu"
repeat_region      21520..21714
                    /rpt_family="B4"
repeat_region      21932..22139
                    /rpt_family="B2"
repeat_region      22457..22672
                    /rpt_family="B4"

```

Query Match 88.2%; Score 15; DB 10; Length 180516;
Best Local Similarity 100.0%; Pred. No. 9.5e+02;

```

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCCGCTAGTAC 15
    |||||
Db 25935 ACCCGCTAGTAC 25921

RESULT 11
ATHS31217/c 245 bp DNA linear PLN 29-MAR-2003
LOCUS Arabidopsis thaliana T-DNA flanking sequence, left border, clone
228D07.
ACCESSION AJ531217
VERSION AJ531217.1 GI:26799477
KEYWORDS left border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F.,
          Chauvin, S., Bechtold, N., Cruaud, C., Derose, R., Pelletier, G.,
          Lepoint, L., Caboche, M. and Lecharny, A.
          T-DNA integration into the Arabidopsis genome depends on sequences
          of pre-insertion sites
          EMBO Rep. 3 (12), 1152-1157 (2002)
JOURNAL MEDLINE
          22363535
PUBMED 12446565
REFERENCE 2 (bases 1 to 245)
AUTHORS Balzerque, S.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue
          Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana
          plants from INRA (Versailles). The DNA fragment(s) resulting from
          the PCR were directly sequenced from the left or the right border
          to determine the genomic sequence flanking the insertion. T-DNA
          derived sequences were removed. Information to order the
          corresponding mutant line and a link to a database providing a
          graphical display of the insertion site are available at
          http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
          been generated in the framework of the French plant genomics
          program 'Genoplante' (http://www.genoplante.com and
          http://genoplante-info.infbio.gen.fr).
FEATURES
    source
        1..245
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"
            /cultivar="Massillaewek1ja"
            /db_xref="taxon:3702"
            /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
            /note="T-DNA flanking sequence
            left border"
    misc_feature
        1..245
            /note="T-DNA flanking sequence
            left border"
ORIGIN
Query Match 84.7%; Score 14.4; DB 8; Length 245;
Best Local Similarity 93.8%; Pred. No. 8.8e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 GCCGCTAGTACT 17
    |||||
Db 70 GCCGCTAGTACT 55

RESULT 12
CQ738165/c 495 bp DNA linear PAT 03-FEB-2004
LOCUS Sequence 24099 from Patent WO02068579.
DEFINITION CQ738165
ACCESSION CQ738165.1 GI:42338344
VERSION

```


TITLE
JOURNAL

Meyer, T., Tsapin, A., Scott, J., Beanan, M., Brinkac, L., Daugherty, S.,
DeBoy, R., Dodson, R., Durkin, A., Haft, D., Kolonay, J., Madupu, R.,
Peterson, J., Umayam, L., White, O., Wolf, A., Yaman, J.,
Weldman, J., Impraim, M., Lee, K., Berry, K., Lee, C., Mueller, J.,
Khoury, H., Gill, J., Uterback, T., McDonald, L., Feldblyum, T.,
Smith, H., Venter, J., Neale, K., and Fraser, C.
Submitted (12-SEP-2002) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA

FEATURES

Source

Location/Qualifiers

1. .10323

/organism="Shewanella oneidensis MR-1"

/mol_type="Genomic DNA"

/strain="MR-1"

/db_xref="taxon:211586"

124. .597

/gene="secB"

/locus_tag="SO0052"

/note="Identified by match to TIGR protein family HMM

TIGR00809"

/codon_start=1

/product="protein-export protein SecB"

/protein_id="AN53139.1"

/db_xref="GI:24345418"

/translation="MAEVANNEQAPQFNIRVYTKDVSETPNSPAVPOKEMNEVK
LDLDRSAKLADVYVLSLTVYQNGSTAFCEVOAGISFISAGLTPOLAHSG
AYCPNLPYAREAVSGSLVGRGTFPQNLAPVNDALFAQVQORQAANAEANA"

602. .1618

/gene="gpaA"

/locus_tag="SO0053"

602. .1618

/locus_tag="SO0053"

/note="similar to GB:L06148, SP:008378, PID:553309,
GB:L06148, SP:008378, and PID:553309; identified by
sequence similarity; putative"

/codon_start=1

/translation="table=1

/product="glycerol-3-phosphate dehydrogenase (NAD(P)+)"

/protein_id="AN53140.1"

/db_xref="GI:24345419"

/translation="MNSADITVVGAGSYTALALISLNSGHTLTMGNDPAMQTLA
EDCNQNFLEGLAEPFECILHEDLAKALASNNVLVVPSPHVGTVLAQKPLRQDA
RIWATKGLPEPRLQDYARDVLEQYPLAVLSGTFPAKELAMGLPTISVAGTDP
QFAELVELHSPKRLVYANDDFIGLQGAIVNVAIAGMSDGIQFANARTALI
TRGLVELTRLGALGASTATFPMAGLIGDVLCTDNOSNRBRFGALGCDVDTAO
VEIGOVVEGRNKEVFTTLAKRLGVEPRITEQIYQVLYQKSPVDAAKELIGREKISE
TPQ"

1968. .3158

/locus_tag="SO0054"

1968. .3158

/locus_tag="SO0054"

/note="Identified by match to Pfam protein family HMM

pf03486"

/codon_start=1

/translation="table=1

/product="conserved hypothetical protein TIGR00275"

/protein_id="AN53141.1"

/db_xref="GI:24345420"

/translation="MSYKHHVITIGGAAGLGCATAGYRGDVLVDNAKQAGRI
LISGGRCNFTNLKVEPANFTGPNPHVNSALARYSQFTELVERHGLEYHNDQ
LPCNDKAKELVIMLLECEWAGVSIKRLDILSVGNQARFELNLSNGSLSCDSVIT
ATGSLMPKGLATPYGOLAEQGLKVLPTHAGLVPTTMSEDIKPEEPGLGAVPSR
ITAKDGTAFSEALPTFHRLGSGPAILOISNWKAGETIKINLLPNMDAQAISQOLAA
HPKQSLRNTLSMLPKRLVEVLFDEVILANKALNOLVHABRAKLVLDLHPVTVMNGE
GYTAEVTLGGVDTHELSSKTMALIKVPGLFPIGEVNDVSGMLGGRFPOAMASGVAA
GKAV"

complement (3381. .5060)

/locus_tag="SO0055"

gene

CDS

complement (3381. .5060)

/locus_tag="SO0055"

/note="similar to SP:P26167, and PID:995679; identified
by sequence similarity; putative"

/codon_start=1

/translation="table=1

/product="conserved domain protein"

/protein_id="AN53142.1"

/db_xref="GI:24345421"

/translation="MRKARDNRPVLTFTGAGSKSAGIPLASELVNEINEKALDLK
ALSTDRYDGRCAALTKDDRRELLQDYLOKARINNAHIVQALLLEKGYORVLTEN
FDNLSRSCGLGIPSIYDFTSANLDELILIDPSIVHLHGSHGFQVOLTNETK
HAELSGFVANTLNQSPSIFIGYSGNADAPPLIKDKFTPOKRLFWGRSKAAHVN
ENLJAPNVAVHVLHNDADRFELIAQELSCPPKARDDPYGHLLDLLEPVLSEPTQO
HNELDILASTRSHMRKQSEEREDARFELMLKGTNELIALKGTLEDEKAAV
AMABIALGDLLEQAKKAQPKLFEOSFEKYQALALIKPMHEKALNMGALSDLAOL
KQEPALFEQSLBKXQALALIKPMHDLFPWGVNLSLAOLKQEPALFEOSIEKYEA
LAIKPDHEAISMWGCALDLAKLQOQTELFEESSKSLIOARSVANOPNVLAQVSL
LHDDLNCKEELPACVYKAKTLPDKNHLNMTDTDLDPVRNKEPFKELDLTPG"

5448. .6206

/locus_tag="SO0056"

5448. .6206

/locus_tag="SO0056"

/note="Identified by match to TIGR protein family HMM

TIGR01096"

/codon_start=1

/translation="table=1

/product="transporter, putative"

/protein_id="AN53143.1"

/db_xref="GI:24345422"

/translation="WVTKPLKMLICLAGASPLAEPVITGABDDWAPYCALDKSG
OPGLAPELVAVPATEKLEIARPPFPACMDASGKVGCPNATITENRQYTH
HKTMPBEDLAFALASBPNDLQSLSEKRGVITGLVPTDPMNTVITREBAS
DAQILELVGRVDYILMNGMPYLOIQOGLGKYIKVGKISTDGFMAFSRTHPG
BAMSKQFEBGLQIKNNGYBELIRDFEAKLGHR"

complement (6318. .7679)

/locus_tag="SO0057"

complement (6318. .7679)

/locus_tag="SO0057"

/note="similar to GP:10173212, and GP:3395637; identified
by sequence similarity; putative"

/codon_start=1

/translation="table=1

/product="potassium uptake protein, Trk family"

/protein_id="AN53144.1"

/db_xref="GI:24345423"

/translation="WVQHPSPLEQOTOKAGKLPQAPFILLSFALLILVGTCLIK
LPATLTPISMLQSLFTVSADVGLVVDGTFTFGQVIVALLIQCGLMTF
AIVTLAIGKIGFLQQTVAKEAFNOTDSTIVSTAKAVIPSLVAVGMILLISVTH
SDPLGMQSLSFHFPTTISAPNNAGSALSDSLMPYVADPVNVLTTTGLFTVGLGFS
VWDLNRHKMSKLTIVSRMTITGITLINNALIAIYLIBEKNPNTLAPLSELGKWA
SWQAVYPTAGFNTLPIKLEDSALLILVLMFICGGSISTASGIRKVPFVLLAT
VSYLRDEAAYAFKREI PKETISKALATWISIGVTWMAIFALVTEKAPILDIAPFA
VSAFVTGLSRGLTSLNSGAGQIIIFMYMGRGLPLMALVPLANPRVKKLRYAEFL
ATG"

complement (7681. .8379)

/locus_tag="SO0058"

complement (7681. .8379)

/locus_tag="SO0058"

/note="similar to GP:10173212, and GP:3288678; identified
by sequence similarity; putative"

/codon_start=1

/translation="table=1

/product="potassium uptake protein, putative"

/protein_id="AN53145.1"

/db_xref="GI:24345424"

/translation="MAHPTVIGVGRGVAVSLLEIHLGHTVTVGSDHKAVERKEVYL
TVAIVDCDADEALRELDASSGVVVAIGDMOSLLCTIALKNDVQVITKASNK
AHTTISKUGAVRIIHPEDMGKRAVQASINPYNNFLAIGKLYIVEIHKHNLQV
TVGLLGLSLHNDASNEQTVNVRNPKGVAPLWAKRLTVPSKIDSEFILTDEALPLC
GSAELGKLAPRLV"

complement (8425. .9114)

/gene="kdpG"

gene

CDS /locus_tag="S00059"
complement(8425..9114)
/gene="KdpE"

/locus_tag="S00059"
/note="similar to SP:002940, and PTD:151433; identified
by sequence similarity; putative"

/codon_start=1
/transl_table=11
/product="transcriptional regulatory protein KdpE"
/protein_id="AAN53146.1"
/db_xref="GI:24345425"
/translation="MAVKYLVVDDPQIHFMRLSABGFYLSATSIGSLKQYRS
HQPVLVDLGLPDGDGIELLMRQDPTPLVLTARPEEKIRILEAGANDYLSK
PGIRLEIVRIKRLDVLDTISIAADVLOFPLKQSSHCMTANQOIRALTKEPAL
LAHLSHPQQLVKQSELLRRIWGETHQEDGHYLRILVSQLRKLNDSDEQULKTEPG
LGYRLTDLGLP"

gene complement(9115..10161)
/locus_tag="S00060"
complement(9115..10161)

CDS /locus_tag="S00060"
/note="similar to GP:9947604; identified by sequence
similarity; putative"

/codon_start=1
/transl_table=11
/product="sensor histidine kinase"
/protein_id="AAN53147.1"
/db_xref="GI:24345426"
/translation="MTFANTASINHYMRTHPALFTAMLVVLITSYFIDVFGSAIA

Query Match 84.7%; Score 14.4; DB 1; Length 10323;
Best Local Similarity 93.8%; Pred. No. 3.8e+03;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACC 16
|||||
Db 2848 AGCCGCTTAGTAACC 2833

Search completed: November 7, 2005, 06:42:15
Job time : 676.064 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 23:04:21 ; Search time 196.043 Seconds
(without alignments)
513.336 Million cell updates/sec

Title: US-10-777-131a-5

Perfect score: 17

Sequence: 1 agcgcgcgtactact 17

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001s:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	100.0	17	8	ABZ79931 Mycobacte
2	17	100.0	2488	8	ABZ79947 Mycobacte
3	17	100.0	110000	4	AA199682_14
4	17	100.0	110000	4	AA199683_14
5	14.4	84.7	505	2	AAZ21192
6	14.4	84.7	735	8	ACA54363
7	14.4	84.7	2967	5	AAS93164
8	14.4	82.4	237	9	ACL21756
9	14.4	82.4	317	9	ACL21751
10	14.4	82.4	360	13	ADR91508
11	14.4	82.4	394	9	ACL21733
12	14.4	82.4	353	9	ACL21749
13	14.4	82.4	561	9	ACL21732
14	14.4	82.4	582	9	ACL21772
15	14.4	82.4	591	9	ACL21729
16	14.4	82.4	596	9	ACL21768
17	14.4	82.4	602	9	ACL21752
18	14.4	82.4	603	9	ACL21770
19	14.4	82.4	605	9	ACL21750
20	14.4	82.4	634	9	ACL21736

21	14	82.4	639	9	ACL21731	ACL21731 DNA clone
22	14	82.4	653	9	ACL21730	ACL21730 DNA clone
23	14	82.4	664	9	ACL21744	ACL21744 DNA clone
24	14	82.4	694	9	ACL21771	ACL21771 DNA clone
25	14	82.4	698	9	ACL21757	ACL21757 DNA clone
26	14	82.4	700	9	ACL21755	ACL21755 DNA clone
27	14	82.4	700	9	ACL21764	ACL21764 DNA clone
28	14	82.4	702	9	ACL21747	ACL21747 DNA clone
29	14	82.4	703	9	ACL21748	ACL21748 DNA clone
30	14	82.4	738	9	ACL21746	ACL21746 DNA clone
31	14	82.4	751	9	ACL21769	ACL21769 DNA clone
32	14	82.4	1299	4	AAF90009	AAF90009 Nucleotid
33	14	82.4	1334	6	AB571611	AB571611 Streptoco
34	14	82.4	1335	10	ADB61679	ADB61679 16S rRNA
35	14	82.4	1385	2	AAV27049	AAV27049 Mouse neu
36	14	82.4	1385	3	AAZ51980	AAZ51980 Murine ne
37	14	82.4	1385	5	AAF27269	AAF27269 Mouse neu
38	14	82.4	1385	9	AAI62216	AAI62216 Mouse ato
39	14	82.4	1470	4	AAF31086	AAF31086 Bacterial
40	14	82.4	1515	5	AA511029	AA511029 Streptoco
41	14	82.4	1515	10	ADB61678	ADB61678 16S rRNA
42	14	82.4	1522	2	AA57721	AA57721 16S rRibos
43	14	82.4	1551	10	ADB61677	ADB61677 16S rRNA
44	14	82.4	6109	4	AAH78667	AAH78667 Murine Co
45	14	82.4	7250	6	ABN72583	ABN72583 Streptoco

ALIGNMENTS

RESULT 1
ID ABZ79931 standard; DNA; 17 BP.
AC ABZ79931;
XX
DT 19-MAY-2003 (first entry)
XX
DE Mycobacterium tuberculosis alka-1 PCR primer SEQ ID NO:5.
XX
KW Mycobacterium tuberculosis; mutr2; alka; ogt; rv3908; muty; rv3909;
KM detection; multidrug resistance; multiple drug resistance; MDR;
KW infection; PCR primer; ss.
XX
OS Mycobacterium tuberculosis.
OS Synthetic.
PN WO2003016562-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002MO-EP009679.
XX
PR 14-AUG-2001; 2001US-0311824P.
PR 21-AUG-2001; 2001US-0313523P.
XX
PA (INSP) INST PASTEUR.
XX
PI Gicquel B;
XX
DR WPI; 2003-256711/25.
XX
PT Predicting the epidemic character of a Mycobacterium tuberculosis isolate
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate
PT by detecting an alteration in the DNA repair system of the isolate.
XX
XX Claim 32; Page 16; 83pp; English.
XX
CC The present invention describes a method for predicting the epidemic
CC character of a Mycobacterium tuberculosis isolate and/or a selective
CC advantage to be maintained in the host and/or the acquisition of multiple
CC drug resistance (MDR) by the isolate comprising detecting an alteration
CC in the DNA repair system of the isolate. Also described: (1) detecting a

CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a
CC polymucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)
CC an *Escherichia coli* strain containing the plasmid pMYC2501; and (5)
CC detecting in a patient infected by Mycobacterium tuberculosis a higher
CC risk of being unable to eliminate the bacillus or of developing MDR
CC tuberculosis. The method is useful for predicting the epidemic character
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to
CC be maintained in the host and/or the acquisition of MDR by the isolate.
CC The present sequence represents a PCR primer for *M. tuberculosis* alkA,
CC which is used in the exemplification of the present invention
SQ Sequence 17 BP; 4 A; 5 C; 5 G; 3 T; 0 U; 0 Other;

Query Match 100.0%; Score 17; DB 8; Length 17;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCGCGTAGTAACCT 17
DB 1 AGCGCGTAGTAACCT 17

RESULT 2
ABZ79947 standard; DNA; 2488 BP.
XX ABZ79947;
XX
XX
XX 19-MAY-2003 (first entry)
XX
XX Mycobacterium tuberculosis alkA nucleotide sequence SEQ ID NO:17.
XX
XX Mycobacterium tuberculosis; mutT2; alkA; ogc; R33908; mutY; R33909;
XX detection; multidrug resistance; multiple drug resistance; MDR;
XX infection; gene; de.
XX
XX Mycobacterium tuberculosis.
XX
XX WO2003016562-A2.
XX
XX 27-FEB-2003.
XX
XX 14-AUG-2002; 2002MO-EP009679.
XX
XX 14-AUG-2001; 2001US-0311824P.
XX 21-AUG-2001; 2001US-0313523P.
XX
XX (INSP) INST PASTEUR.
XX
XX Gicquel B;
XX
XX WPI; 2003-256711/25.
XX
XX Predicting the epidemic character of a Mycobacterium tuberculosis isolate
XX and/or the acquisition of multiple drug resistance (MDR) by the isolate
XX by detecting an alteration in the DNA repair system of the isolate.
XX
XX Disclosure; Fig 2A-B; 83pp; English.
XX
XX The present invention describes a method for predicting the epidemic
XX character of a Mycobacterium tuberculosis isolate and/or a selective
XX advantage to be maintained in the host and/or the acquisition of multiple
XX drug resistance (MDR) by the isolate comprising detecting (1) detecting a
XX in the DNA repair system of the isolate. Also described: (1) detecting a
XX Mycobacterium tuberculosis strain with a MDR phenotype; (2) a
XX polymucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)
XX an *Escherichia coli* strain containing the plasmid pMYC2501; and (5)
XX detecting in a patient infected by Mycobacterium tuberculosis a higher
XX risk of being unable to eliminate the bacillus or of developing MDR
XX tuberculosis. The method is useful for predicting the epidemic character
XX of a Mycobacterium tuberculosis isolate and/or a selective advantage to
XX be maintained in the host and/or the acquisition of MDR by the isolate.
XX The present sequence represents a *M. tuberculosis* alkA nucleotide

CC sequence, which is used in the exemplification of the present invention
XX SQ Sequence 2488 BP; 426 A; 847 C; 807 G; 408 T; 0 U; 0 Other;

Query Match 100.0%; Score 17; DB 8; Length 2488;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCGCGTAGTAACCT 17
DB 407 AGCGCGTAGTAACCT 423

RESULT 3
AA199682_14/C
Continuation (15 of 45) of AA199682 from base 1400001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments
WP Fragment Name Begin End
WP AA199682_00 1 110000
WP AA199682_01 100001 210000
WP AA199682_02 200001 310000
WP AA199682_03 300001 410000
WP AA199682_04 400001 510000
WP AA199682_05 500001 610000
WP AA199682_06 600001 710000
WP AA199682_07 700001 810000
WP AA199682_08 800001 910000
WP AA199682_09 900001 1010000
WP AA199682_10 1000001 1110000
WP AA199682_11 1100001 1210000
WP AA199682_12 1200001 1310000
WP AA199682_13 1300001 1410000
WP AA199682_14 1400001 1510000
WP AA199682_15 1500001 1610000
WP AA199682_16 1600001 1710000
WP AA199682_17 1700001 1810000
WP AA199682_18 1800001 1910000
WP AA199682_19 1900001 2010000
WP AA199682_20 2000001 2110000
WP AA199682_21 2100001 2210000
WP AA199682_22 2200001 2310000
WP AA199682_23 2300001 2410000
WP AA199682_24 2400001 2510000
WP AA199682_25 2500001 2610000
WP AA199682_26 2600001 2710000
WP AA199682_27 2700001 2810000
WP AA199682_28 2800001 2910000
WP AA199682_29 2900001 3010000
WP AA199682_30 3000001 3110000
WP AA199682_31 3100001 3210000
WP AA199682_32 3200001 3310000
WP AA199682_33 3300001 3410000
WP AA199682_34 3400001 3510000
WP AA199682_35 3500001 3610000
WP AA199682_36 3600001 3710000
WP AA199682_37 3700001 3810000
WP AA199682_38 3800001 3910000
WP AA199682_39 3900001 4010000
WP AA199682_40 4000001 4110000
WP AA199682_41 4100001 4210000
WP AA199682_42 4200001 4310000
WP AA199682_43 4300001 4410000
WP AA199682_44 4400001 4411529

Query Match 100.0%; Score 17; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCGCGTAGTAACCT 17
DB 79210 AGCGCGTAGTAACCT 79194

XX New antisense nucleic acid, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
PS Claim 14; SEQ ID NO 42233; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway; (8)
CC required for proliferation, or that inhibits cellular proliferation; (9)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 735 BP; 180 A; 141 C; 214 G; 200 T; 0 U; 0 Other;
XX
Query Match 84.7%; Score 14.4; DB 8; Length 735;
Best Local Similarity 93.8%; Pred. No. 3.1e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 GCGCGTAGGTAACCT 17
DB 458 GCGCGTAGGTAACCT 443
XX
RESULT 7
AAS93164/c
ID AAS93164 standard; cDNA; 2967 BP.
XX
AC AAS93164;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #28968.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX

XX
PA (HXS-) HXSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX MPI; 2001-639362/73.
XX
XX P-PSDB; ABG28977.
XX
PS Claim 1; SEQ ID NO 28968; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2967 BP; 689 A; 782 C; 772 G; 724 T; 0 U; 0 Other;
XX
Query Match 84.7%; Score 14.4; DB 5; Length 2967;
Best Local Similarity 93.8%; Pred. No. 3.3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 GCGCGTAGGTAACCT 17
DB 2267 GCGCGTAGGTAACAT 2252
XX
RESULT 8
ACT21756
ID ACT21756 standard; DNA; 237 BP.
XX
AC ACT21756;
XX
DT 27-OCT-2003 (revised)
XX
DT 17-OCT-2003 (first entry)
XX
DE DNA clone originating in barley containing SNP encoding sequence #11747.
XX
XX Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KM gene; ss.
XX
XX Hordeum vulgare; ssp. spontaneum.
XX
XX WO2003057877-A1.
XX
XX 17-JUL-2003.
XX
XX 16-DEC-2002; 2002WO-IB005403.
XX
XX 20-DEC-2001; 2001JP-00387059.
XX
XX 20-DEC-2001; 2001JP-00387131.
XX
XX 20-DEC-2001; 2001JP-00403299.
XX
XX

PR 20-DEC-2001; 2001JP-00403300.
PR 27-SEP-2002; 2002JP-00327515.
XX
PA (UYN1-) UNIV JAPAN OKAYAMA.
XX
PI Sato K, Takeda K, Kohara Y;
XX
DR MPI; 2003-587127/55.
XX
PT Single nucleotide polymorphism sites in barley varieties and DNA
PT sequences containing them for analysis and identification of barley
PT varieties and production of barley transformants with desired
PT characteristics.
XX
PS Disclosure; SEQ ID XX; 284bp; Japanese.
XX
CC The present invention relates to oligonucleotide clones originating in
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
CC varieties, identification of particular varieties and genotype-phenotype
CC analysis, isolation of specific genes and creation of new varieties by
CC transformation of barley varieties with them and production of new barley
CC varieties with desired properties. The present sequence represents an
CC oligonucleotide clone DNA sequence featured in the specification. The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 237 BP; 33 A; 80 C; 64 G; 59 T; 0 U; 1 Other;
XX
Query Match 82.4%; Score 14; DB 9; Length 237;
Best Local Similarity 100.0%; Pred. No. 5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GCCGCGTAGTAAC 15
DB 88 GCCGCGTAGTAAC 101
XX
RESULT 9
ACL21751
ID ACL21751 standard; DNA; 317 BP.
XX
AC ACL21751;
XX
DT 27-OCT-2003 (revised)
DT 17-OCT-2003 (first entry)
XX
DE DNA clone originating in barley containing SNP encoding sequence #11742.
XX
KM Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KM gene; ss.
XX
OS Hordeum vulgare; ssp. spontaneum.
XX
PN WO2003057877-A1.
XX
PD 17-JUL-2003.
XX
PF 16-DEC-2002; 2002WO-1B005403.
XX
PR 20-DEC-2001; 2001JP-00387059.
PR 20-DEC-2001; 2001JP-00387131.
PR 20-DEC-2001; 2001JP-00403299.
PR 20-DEC-2001; 2001JP-00403300.
PR 27-SEP-2002; 2002JP-00327515.
XX
PA (UYN1-) UNIV JAPAN OKAYAMA.
XX
PI Sato K, Takeda K, Kohara Y;
XX
DR MPI; 2003-587127/55.

XX
PT Single nucleotide polymorphism sites in barley varieties and DNA
PT sequences containing them for analysis and identification of barley
PT varieties and production of barley transformants with desired
PT characteristics.
XX
PS Disclosure; SEQ ID XX; 284bp; Japanese.
XX
CC The present invention relates to oligonucleotide clones originating in
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
CC varieties, identification of particular varieties and genotype-phenotype
CC analysis, isolation of specific genes and creation of new varieties by
CC transformation of barley varieties with them and production of new barley
CC varieties with desired properties. The present sequence represents an
CC oligonucleotide clone DNA sequence featured in the specification. The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 317 BP; 60 A; 95 C; 91 G; 71 T; 0 U; 0 Other;
XX
Query Match 82.4%; Score 14; DB 9; Length 317;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GCCGCGTAGTAAC 15
DB 268 GCCGCGTAGTAAC 281
XX
RESULT 10
ADR91508/c
ID ADR91508 standard; DNA; 360 BP.
XX
AC ADR91508;
XX
DT 16-DEC-2004 (first entry)
XX
DE Novel S. pneumoniae DNA sequence, SEQ ID 143.
XX
DE Meningitis; bacteraemia; pneumonia; otitis media; ds;
KM bacterial infection.
XX
OS Streptococcus pneumoniae.
XX
PN US6800744-B1.
XX
PD 05-OCT-2004.
XX
PF 30-JUN-1998; 98US-00107433.
XX
PR 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR MPI; 2004-697205/68.
DR P-PsDB; ADR94111.
XX
PT New isolated nucleic acid encoding a Streptococcus pneumoniae
PT polypeptide, useful for diagnosing, preventing and/or treating
PT pathological conditions resulting from the bacterial infection.
XX
PS Disclosure; SEQ ID NO 143; 151bp; English.
XX
CC The invention relates to an isolated nucleic acid comprising a sequence
CC encoding a Streptococcus pneumoniae ADR9136polypeptide, or its
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,

CC ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
CC hybridizable under high stringency conditions to the nucleotide sequence.
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.
CC Also included are a recombinant expression vector comprising the isolated
CC nucleic acid cited above operably linked to a transcription regulatory
CC element, a cell comprising the recombinant expression vector and a probe
CC comprising at least 20 consecutive nucleotides of the nucleotide
CC sequences as cited above. The methods and compositions of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC pathological conditions resulting from bacterial infection by
CC *Streptococcus pneumoniae* e.g. pneumonia, bacteremia, meningitis and
CC otitis media. The present sequence is one of the 2603 disclosed S.
CC *pneumoniae* nucleic acid sequences. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docid=680074481.
CC
SQ Sequence 360 BP; 80 A; 98 C; 59 G; 123 T; 0 U; 0 Other;

Query Match 82.4%; Score 14; DB 13; Length 360;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CCGGTAGTAACT 17
DB 49 CCGGTAGTAACT 36

RESULT 11
ACL21733
XX ACL21733 standard; DNA; 394 BP.
XX
ACL21733;
XX
DT 27-OCT-2003 (revised)
DT 17-OCT-2003 (first entry)
XX
DE DNA clone originating in barley containing SNP encoding sequence #11724.
XX
KM Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KM gene; ss.
XX
OS *Hordeum vulgare*; var. (cul.Haruna Nijo).
XX
PN WO2003057877-A1.
XX
PD 17-JUL-2003.
XX
PF 16-DEC-2002; 2002WO-IB005403.
XX
PR 20-DEC-2001; 2001JP-00387059.
PR 20-DEC-2001; 2001JP-00387131.
PR 20-DEC-2001; 2001JP-00403299.
PR 20-DEC-2001; 2001JP-00403300.
PR 27-SRP-2002; 2002JP-00327515.
XX
PA (UNNT-) UNIV JAPAN OKAYAMA.
XX
PI Sato K, Takeda K, Kohara Y;
XX
DR WPI; 2003-587127/55.
XX
PT Single nucleotide polymorphism sites in barley varieties and DNA
PT sequences containing them for analysis and identification of barley
PT varieties and production of barley transformants with desired
PT characteristics.
XX
PS Disclosure; SEQ ID XX; 284bp; Japanese.
XX
CC The present invention relates to oligonucleotide clones originating in

CC barley (*Hordeum vulgare*) which contain single nucleotide polymorphisms
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
CC varieties, identification of particular varieties and genotype-phenotype
CC analysis, isolation of specific genes and creation of new varieties by
CC transformation of barley varieties with them and production of new barley
CC varieties with desired properties. The present sequence represents an
CC oligonucleotide clone DNA sequence featured in the specification. The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published-pct-sequences. (updated on 27-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 394 BP; 70 A; 121 C; 107 G; 96 T; 0 U; 0 Other;

Query Match 82.4%; Score 14; DB 9; Length 394;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCCGCTAGTAAAC 15
DB 260 GCCGCTAGTAAAC 273

RESULT 12
ACL21749
XX ACL21749 standard; DNA; 553 BP.
XX
ACL21749;
XX
DT 27-OCT-2003 (revised)
DT 17-OCT-2003 (first entry)
XX
DE DNA clone originating in barley containing SNP encoding sequence #11740.
XX
KM Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KM gene; ss.
XX
OS *Hordeum vulgare*; ssp. spontaneum.
XX
PN WO2003057877-A1.
XX
PD 17-JUL-2003.
XX
PF 16-DEC-2002; 2002WO-IB005403.
XX
PR 20-DEC-2001; 2001JP-00387059.
PR 20-DEC-2001; 2001JP-00387131.
PR 20-DEC-2001; 2001JP-00403299.
PR 20-DEC-2001; 2001JP-00403300.
PR 27-SRP-2002; 2002JP-00327515.
XX
PA (UNNT-) UNIV JAPAN OKAYAMA.
XX
PI Sato K, Takeda K, Kohara Y;
XX
DR WPI; 2003-587127/55.
XX
PT Single nucleotide polymorphism sites in barley varieties and DNA
PT sequences containing them for analysis and identification of barley
PT varieties and production of barley transformants with desired
PT characteristics.
XX
PS Disclosure; SEQ ID XX; 284bp; Japanese.
XX
XX
CC The present invention relates to oligonucleotide clones originating in
CC barley (*Hordeum vulgare*) which contain single nucleotide polymorphisms
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
CC varieties, identification of particular varieties and genotype-phenotype
CC analysis, isolation of specific genes and creation of new varieties by
CC transformation of barley varieties with them and production of new barley
CC varieties with desired properties. The present sequence represents an
CC oligonucleotide clone DNA sequence featured in the specification. The
CC sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
CC standardise OS field)

XX Sequence 553 BP; 134 A; 165 C; 136 G; 114 T; 0 U; 4 Other;

Query Match 82.4%; Score 14; DB 9; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCGCGGTAGGTAC 15
DB 513 GCGCGGTAGGTAC 526

RESULT 13

ACL21732
ID ACL21732 standard; DNA; 561 BP.

XX ACL21732;

DT 27-OCT-2003 (revised)
DT 17-OCT-2003 (first entry)

DE DNA clone originating in barley containing SNP encoding sequence #11723.

KM Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KM gene; ss.

OS Hordeum vulgare; var. (cul.Haruna Nijo).

PN WO2003057877-A1.

XX 17-JUL-2003.

PD 16-DEC-2002; 2002WO-IB005403.

XX 20-DEC-2001; 2001JP-00387059.

PR 20-DEC-2001; 2001JP-00387131.

PR 20-DEC-2001; 2001JP-00403299.

PR 27-SEP-2002; 2002JP-00327515.

XX (UYNI-) UNIV JAPAN OKAYAMA.

XX Sato K, Takeda K, Kohara Y;

XX WPI; 2003-587127/55.

PT Single nucleotide polymorphism sites in barley varieties and DNA
PT sequences containing them for analysis and identification of barley
PT varieties and production of barley transformants with desired
PT characteristics.

PS Disclosure; SEQ ID XX; 284pp; Japanese.

XX The present invention relates to oligonucleotide clones originating in
XX barley (Hordeum vulgare) which contain single nucleotide polymorphisms
XX (SNP). The oligonucleotides may be used for analysis of SNPs among barley
XX varieties, identification of particular varieties and genotype-phenotype
XX analysis, isolation of specific genes and creation of new varieties by
XX transformation of barley varieties with them and production of new barley
XX varieties with desired properties. The present sequence represents an
XX oligonucleotide clone DNA sequence featured in the specification. The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
XX standardise OS field)

XX Sequence 561 BP; 111 A; 166 C; 148 G; 136 T; 0 U; 0 Other;

Query Match 82.4%; Score 14; DB 9; Length 561;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCGCGGTAGGTAC 15

DB 415 GCGCGGTAGGTAC 428

RESULT 14

ACL21772
ID ACL21772 standard; DNA; 582 BP.

XX ACL21772;

DT 27-OCT-2003 (revised)
DT 17-OCT-2003 (first entry)

DE DNA clone originating in barley containing SNP encoding sequence #11763.

KM Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KM gene; ss.

OS Hordeum vulgare; ssp. spontaneum.

PN WO2003057877-A1.

XX 17-JUL-2003.

PD 16-DEC-2002; 2002WO-IB005403.

XX 20-DEC-2001; 2001JP-00387059.

PR 20-DEC-2001; 2001JP-00387131.

PR 20-DEC-2001; 2001JP-00403299.

PR 27-SEP-2002; 2002JP-00327515.

XX (UYNI-) UNIV JAPAN OKAYAMA.

XX Sato K, Takeda K, Kohara Y;

XX WPI; 2003-587127/55.

PT Single nucleotide polymorphism sites in barley varieties and DNA
PT sequences containing them for analysis and identification of barley
PT varieties and production of barley transformants with desired
PT characteristics.

PS Disclosure; SEQ ID XX; 284pp; Japanese.

XX The present invention relates to oligonucleotide clones originating in
XX barley (Hordeum vulgare) which contain single nucleotide polymorphisms
XX (SNP). The oligonucleotides may be used for analysis of SNPs among barley
XX varieties, identification of particular varieties and genotype-phenotype
XX analysis, isolation of specific genes and creation of new varieties by
XX transformation of barley varieties with them and production of new barley
XX varieties with desired properties. The present sequence represents an
XX oligonucleotide clone DNA sequence featured in the specification. The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
XX standardise OS field)

XX Sequence 582 BP; 139 A; 174 C; 140 G; 124 T; 0 U; 5 Other;

Query Match 82.4%; Score 14; DB 9; Length 582;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCGCGGTAGGTAC 15
DB 521 GCGCGGTAGGTAC 534

RESULT 15

ACL21729
 ID ACL21729 standard; DNA; 591 BP.
 XX
 AC ACL21729;
 XX
 DT 27-OCT-2003 (revised)
 DT 17-OCT-2003 (first entry)
 XX
 DE DNA clone originating in barley containing SNP encoding sequence #11720.
 XX
 KM Barley, single nucleotide polymorphism; SNP; genotype-phenotype analysis;
 KM gene; ss.
 XX
 OS Hordeum vulgare; var. (cul.Haruna Nijo).
 XX
 PN WO2003057877-A1.
 XX
 PD 17-JUL-2003.
 XX
 PF 16-DEC-2002; 2002WO-1B005403.
 XX
 PR 20-DEC-2001; 2001JP-00387059.
 PR 20-DEC-2001; 2001JP-00387131.
 PR 20-DEC-2001; 2001JP-00403299.
 PR 20-DEC-2001; 2001JP-00403300.
 PR 27-SRP-2002; 2002JP-00327515.
 XX
 PA (UYN1-) UNIV JAPAN OKAYAMA.
 XX
 PI Sato K, Takeda K, Kohara Y;
 XX
 DR MPI; 2003-587127/55.
 XX
 PT Single nucleotide polymorphism sites in barley varieties and DNA
 PT sequences containing them for analysis and identification of barley
 PT varieties and production of barley transformants with desired
 PT characteristics.
 XX
 PS Disclosure; SEQ ID XX; 284bp; Japanese.
 XX
 CC The present invention relates to oligonucleotide clones originating in
 CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
 CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
 CC varieties, identification of particular varieties and genotype-phenotype
 CC analysis, isolation of specific genes and creation of new varieties by
 CC transformation of barley varieties with them and production of new barley
 CC varieties with desired properties. The present sequence represents an
 CC oligonucleotide clone DNA sequence featured in the specification. The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 591 BP; 103 A; 162 C; 139 G; 186 T; 0 U; 1 Other;
 Query Match 82.4%; Score 14; DB 9; Length 591;
 Best Local Similarity 100.0%; Pred. No. 5.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GCCGCGTAGGTAAAC 15
 |||||||||
 Db 60 GCCGCGTAGGTAAAC 73

Search completed: November 7, 2005, 05:40:40
 Job time : 203.043 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2005, 20:56:42 ; Search time 423.915 Seconds
(without alignments)
331.640 Million cell updates/sec

Title: US-10-777-131A-5

Perfect score: 17
Sequence: 1 agccgcgtagtaacct 17

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
23: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
24: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
25: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
26: /cgn2_6/ptodata/2/pubpna/US11A_NEW_PUB.seq.*
27: /cgn2_6/ptodata/2/pubpna/US11A_NEW_PUB.seq.*
28: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	100.0	17	US-10-216-817-5	Sequence 5, Appli
2	17	100.0	17	US-10-777-131A-5	Sequence 5, Appli
3	17	100.0	2488	US-10-216-817-17	Sequence 17, Appli
4	17	100.0	2488	US-10-777-131A-17	Sequence 17, Appli
5	15	88.2	526	US-10-425-115-180069	Sequence 180069,

C	6	14.4	84.7	473	13	US-09-925-065A-480693	Sequence 480693,
C	7	14.4	84.7	735	18	US-10-282-122A-42223	Sequence 42223,
C	8	14.4	84.7	2967	24	US-10-450-763-28968	Sequence 28968,
C	9	14.4	84.7	3186778	14	US-10-027-632-174961	Sequence 174961,
C	10	14.4	84.7	3186778	18	US-10-027-632-174961	Sequence 174961,
C	11	14.4	82.4	360	24	US-10-617-330-143	Sequence 143, App
C	12	14.4	82.4	689	14	US-10-027-632-240523	Sequence 240523,
C	13	14.4	82.4	689	18	US-10-027-632-240523	Sequence 240523,
C	14	14.4	82.4	1334	9	US-09-737-190A-2	Sequence 2, Appli
C	15	14.4	82.4	1335	17	US-10-029-397A-31	Sequence 31, Appli
C	16	14.4	82.4	1385	14	US-10-004-717-30	Sequence 30, Appli
C	17	14.4	82.4	1385	18	US-10-425-259-17	Sequence 17, Appli
C	18	14.4	82.4	1385	21	US-10-860-373-30	Sequence 30, Appli
C	19	14.4	82.4	1385	21	US-10-860-724-30	Sequence 30, Appli
C	20	14.4	82.4	1515	9	US-09-726-774-9	Sequence 9, Appli
C	21	14.4	82.4	1515	17	US-10-029-397A-30	Sequence 30, Appli
C	22	14.4	82.4	1515	20	US-10-719-633-9	Sequence 9, Appli
C	23	14.4	82.4	1522	18	US-10-461-990-1	Sequence 1, App
C	24	14.4	82.4	1551	17	US-10-029-397A-29	Sequence 29, App
C	25	14.4	82.4	6109	10	US-09-795-061-1	Sequence 1, App
C	26	14.4	82.4	21338	8	US-08-961-527-20	Sequence 20, App
C	27	14.4	82.4	21338	18	US-10-158-844-20	Sequence 20, App
C	28	14.4	82.4	2162598	22	US-10-472-928-4979	Sequence 4979,
C	29	14.4	82.4	2162598	22	US-10-472-928-4979	Sequence 4979,
C	30	13.8	81.2	201	20	US-10-741-601-21465	Sequence 21465,
C	31	13.8	81.2	201	22	US-10-741-600-60754	Sequence 60754,
C	32	13.8	81.2	258	21	US-10-425-115-161222	Sequence 161222,
C	33	13.8	81.2	268	21	US-10-425-115-183778	Sequence 183778,
C	34	13.8	81.2	330	19	US-10-424-599-14635	Sequence 14635,
C	35	13.8	81.2	385	21	US-10-425-115-8899	Sequence 8899, A
C	36	13.8	81.2	408	19	US-10-424-599-27014	Sequence 27014,
C	37	13.8	81.2	429	9	US-09-864-761-3181	Sequence 3181, App
C	38	13.8	81.2	455	13	US-09-925-065A-338066	Sequence 338066,
C	39	13.8	81.2	457	13	US-09-925-065A-407156	Sequence 407156,
C	40	13.8	81.2	472	18	US-10-242-535A-30068	Sequence 30068,
C	41	13.8	81.2	473	19	US-10-085-783A-30068	Sequence 30068,
C	42	13.8	81.2	473	19	US-10-424-599-137820	Sequence 137820,
C	43	13.8	81.2	519	21	US-10-425-115-170941	Sequence 170941,
C	44	13.8	81.2	532	14	US-10-027-632-290336	Sequence 290336,
C	45	13.8	81.2	532	14	US-10-027-632-290337	Sequence 290337,

ALIGNMENTS

RESULT 1
US-10-216-817-5
; Sequence 5, Application US/10216817
; Publication No. US20030129619A1
GENERAL INFORMATION:
APPLICANT: GICQUEL, BRIGITTE
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
FILE REFERENCE: 03495.0233-00000
CURRENT APPLICATION NUMBER: US/10/216,817
PRIOR FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: 60/311,824
PRIOR FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: 60/313,523
PRIOR FILING DATE: 2001-08-21
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 17
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Description of Artificial Sequence: Primer
US-10-216-817-5
Query Match 100.0%; Score 17; DB 16; Length 17;
Best Local Similarity 100.0%; Pred. No. 7.3;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACCT 17
|||||
Db 1 AGCCGCGTAGTAACCT 17

RESULT 2
US-10-777-131A-5
; Sequence 5, Application US/10777131A
; Publication No. US20050026216A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; FILE REFERENCE: 02356.0090-00000
; CURRENT APPLICATION NUMBER: US/10/777,131A
; CURRENT FILING DATE: 2004-02-13
; PRIOR APPLICATION NUMBER: PCT/EP02/09679
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 5
; LENGTH: 17
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer
US-10-777-131A-5

Query Match 100.0%; Score 17; DB 22; Length 17;
Best Local Similarity 100.0%; Pred. No. 7.3; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACCT 17
|||||
Db 1 AGCCGCGTAGTAACCT 17

RESULT 3
US-10-216-817-17
; Sequence 17, Application US/10216817
; Publication No. US20030129619A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; FILE REFERENCE: 03495.0233-00000
; CURRENT APPLICATION NUMBER: US/10/216,817
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 2488
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-216-817-17

Query Match 100.0%; Score 17; DB 16; Length 2488;
Best Local Similarity 100.0%; Pred. No. 7.6; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACCT 17

Db 407 AGCCGCGTAGTAACCT 423
|||||

RESULT 4
US-10-777-131A-17
; Sequence 17, Application US/10777131A
; Publication No. US20050026216A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; FILE REFERENCE: 02356.0090-00000
; CURRENT APPLICATION NUMBER: US/10/777,131A
; CURRENT FILING DATE: 2004-02-13
; PRIOR APPLICATION NUMBER: PCT/EP02/09679
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 17
; LENGTH: 2488
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-777-131A-17

Query Match 100.0%; Score 17; DB 22; Length 2488;
Best Local Similarity 100.0%; Pred. No. 7.6; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACCT 17
|||||
Db 407 AGCCGCGTAGTAACCT 423

RESULT 5
US-10-425-115-180069/c
; Sequence 180069, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 180069
; LENGTH: 526
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(526)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577_95802C.1
US-10-425-115-180069

Query Match 88.2%; Score 15; DB 21; Length 526;
Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCCGCGTAGTAAC 16
|||||
Db 127 GCCGCGTAGTAAC 113

```
RESULT 6
US-09-925-065A-480693/C
; Sequence 480693, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 480693
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-480693

Query Match
Best Local Similarity 84.7%; Score 14.4; DB 13; Length 473;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCCGCGTAGTAACC 16
DB 210 ACCCGCGCAGTAACC 195

RESULT 7
US-10-282-122A-42233/C
; Sequence 42233, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
```

```
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42233
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Yersinia pestis
US-10-282-122A-42233

Query Match
Best Local Similarity 84.7%; Score 14.4; DB 18; Length 735;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCCGCGTAGTAACCT 17
DB 458 GCCGCGTAGTAACCT 443

RESULT 8
US-10-450-763-28968/C
; Sequence 28968, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: HySeq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 28968
; LENGTH: 2967
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (2488)..(2769)
; OTHER INFORMATION: 65% homologous to Salmonella typhimurium L72 92% identity
; OTHER INFORMATION: with E. coli lipopolysaccharide biosynthesis protein (WCEC)
; OTHER INFORMATION: (SP:P27833); contains similarity to Pfam family PF01041,
; OTHER INFORMATION: accession number AF233324, Smith-Waterman Score=293.
US-10-450-763-28968

Query Match
Best Local Similarity 84.7%; Score 14.4; DB 24; Length 2967;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCCGCGTAGTAACCT 17
DB 2267 GCCGCGTAGTAACAT 2252

RESULT 9
US-10-027-632-174961
; Sequence 174961, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 174961
LENGTH: 3186778
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(3186778)
US-10-027-632-174961

Query Match 84.7%; Score 14.4; DB 14; Length 3186778;
Best Local Similarity 93.8%; Pred. No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCGCGTAGTAACCT 17
DB 553211 GCGCGTAGGAACCT 553226

RESULT 10
US-10-027-632-174961
Sequence 174961, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 174961
LENGTH: 3186778
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(3186778)

OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174961

Query Match 84.7%; Score 14.4; DB 18; Length 3186778;
Best Local Similarity 93.8%; Pred. No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCGCGTAGTAACCT 17
DB 553211 GCGCGTAGGAACCT 553226

RESULT 11
US-10-617-320-143/C
Sequence 143, Application US/10617320
Publication No. US20050136404A1
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
THERAPEUTICS
NUMBER OF SEQUENCES: 5206
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/617,320
FILING DATE: 10-Jul-2003
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
APPLICATION NUMBER: 60/085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Atinziello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 143:
SEQUENCE CHARACTERISTICS:
LENGTH: 360 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULAR TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...360
SEQUENCE DESCRIPTION: SEQ ID NO: 143:
US-10-617-320-143
Query Match 82.4%; Score 14; DB 24; Length 360;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 49 CGCGTAGTAGTACT 36

RESULT 12
US-10-027-632-240523

Sequence 240523, Application US/10027632
Publication No. US2002019837A1
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 240523

LENGTH: 689

TYPE: DNA

ORGANISM: Human

US-10-027-632-240523

Query Match 82.4%; Score 14; DB 14; Length 689;
Best Local Similarity 87.5%; Pred. No. 4.5e+02;
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGGTAGTAGTACC 16
Db 422 AGCCCGTAGTAGTACC 437

RESULT 13
US-10-027-632-240523

Sequence 240523, Application US/10027632

Publication No. US20030204075A9

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 240523
LENGTH: 689
TYPE: DNA
ORGANISM: Human
US-10-027-632-240523

Query Match 82.4%; Score 14; DB 14; Length 689;
Best Local Similarity 87.5%; Pred. No. 4.5e+02;
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGGTAGTAGTACC 16
Db 422 AGCCCGTAGTAGTACC 437

RESULT 14
US-09-737-190A-2

Sequence 2, Application US/09737190A

Patent No. US20020102545A1

GENERAL INFORMATION:

APPLICANT: Shibuya, Tetsuo

TITLE OF INVENTION: A Method for Changing a Target Array, a Method for Analyzing
a Structure, and an Apparatus, a Storage Medium and a

FILE REFERENCE: JP91990270US1 (14043)

CURRENT APPLICATION NUMBER: US/09/737,190A

PRIOR FILING DATE: 2000-12-14

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 1334

TYPE: DNA

ORGANISM: Streptococcus anginosus

FEATURE:

NAME/KEY: UNSURE

LOCATION: (412)

OTHER INFORMATION: n at position 412 is uncertain

NAME/KEY: UNSURE

LOCATION: (449)

OTHER INFORMATION: n at position 449 is uncertain

NAME/KEY: UNSURE

LOCATION: (701)

OTHER INFORMATION: n at position 701 is uncertain

NAME/KEY: UNSURE

LOCATION: (855)

OTHER INFORMATION: n at position 855 is uncertain

NAME/KEY: UNSURE

LOCATION: (869)

OTHER INFORMATION: n at position 869 is uncertain

NAME/KEY: UNSURE

LOCATION: (909)

OTHER INFORMATION: n at position 909 is uncertain

NAME/KEY: UNSURE

LOCATION: (1018)

OTHER INFORMATION: n at position 1018 is uncertain

NAME/KEY: UNSURE

LOCATION: (1070)

OTHER INFORMATION: n at position 1070 is uncertain

NAME/KEY: UNSURE

LOCATION: (1121)

OTHER INFORMATION: n at position 1121 is uncertain

NAME/KEY: UNSURE

LOCATION: (1228)

OTHER INFORMATION: n at position 1228 is uncertain

NAME/KEY: UNSURE

LOCATION: (1304)

OTHER INFORMATION: n at position 1304 is uncertain

US-09-737-190A-2

Query Match 82.4%; Score 14; DB 9; Length 1334;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 CGCGTAGGTAACCT 17
 |||||
 Db 15 CGCGTAGGTAACCT 28

RESULT 15

US-10-029-397A-31
 ; Sequence 31, Application US/10029397A
 ; Publication No. US20030175709A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MURPHY, GEORGE L.
 ; APPLICANT: WHITLEY, J. PENN
 ; TITLE OF INVENTION: METHOD AND SYSTEM FOR DEPLETING RNA POPULATIONS
 ; FILE REFERENCE: AMB1:076US
 ; CURRENT APPLICATION NUMBER: US/10/029,397A
 ; CURRENT FILING DATE: 2002-03-19
 ; NUMBER OF SEQ ID NOS: 73
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 31
 ; LENGTH: 1335
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pyogenes
 US-10-029-397A-31

Query Match 82.4%; Score 14; DB 17; Length 1335;
 Best Local Similarity 100.0%; Pred.No.4.5e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 CGCGTAGGTAACCT 17
 |||||
 Db 15 CGCGTAGGTAACCT 28

Search completed: November 7, 2005, 05:21:31
 Job time : 440.915 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 02:38:52 ; Search time 1700.72 Seconds
(without alignments)
380.481 Million cell updates/sec

Title: US-10-777-131A-5

Perfect score: 17
Sequence: 1 agccgcgtagtaacct 17

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_esc1:*
2: gb_esc2:*
3: gb_esc3:*
4: gb_esc4:*
5: gb_esc5:*
6: gb_esc6:*
7: gb_esc7:*
8: gb_esc8:*
9: gb_esc9:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	94.1	588	6	CD811742 BM10.001L
2	15.4	90.6	235	7	CV101208 UMC-PCR7D
3	15.4	90.6	411	4	BM539228 h206f02.9
4	15.4	90.6	528	6	CA217548 SCRUD113
5	15.4	90.6	533	7	CV224495 CS_hyp_19
6	15.4	90.6	667	7	CV224494 CS_hyp_19
7	15.4	90.6	700	5	BO294905 WHE2855 G
8	15.4	90.6	768	9	CL653433 PR10118C
9	15	88.2	509	6	CD445870 E101T0204
10	15	88.2	724	9	CNS0607P T3 end of
11	15	88.2	821	8	BZ797610 PUFRCY69TD
12	15	88.2	927	9	CG063514 PUKARS9TD
13	15	88.2	1107	2	BF672870 602152718
14	15	88.2	398	7	CK914339 PJfmgcF 0
15	14.4	84.7	419	1	AJ828384 AJ828384
16	14.4	84.7	472	7	CF604790 RADIC01_0
17	14.4	84.7	480	1	AJ282503 4A3A-AAY-
18	14.4	84.7	487	5	BQ499259 EST08484
19	14.4	84.7	520	5	BP743576 BP743576
20	14.4	84.7	535	5	BO503246 EST12476
21	14.4	84.7	546	7	CF604144 BACCA01.0
22	14.4	84.7	558	8	AZ438618 1M0228K14
23	14.4	84.7	578	5	BX602283 BX602283
24	14.4	84.7	596	4	BM635084 170006875

25	14.4	84.7	608	7	CK708350	CK708350 ZF201-P00
26	14.4	84.7	614	3	BM651965	BM651965 170006873
27	14.4	84.7	622	3	CNS09XEG	EX062516 Single re
28	14.4	84.7	628	5	BM170980	BM170980 BM170980
29	14.4	84.7	639	4	BM613376	BM613376 170006591
30	14.4	84.7	647	4	BM602408	BM602408 170006870
31	14.4	84.7	649	4	BM589600	BM589600 170006884
32	14.4	84.7	651	5	BX603222	BX603222 BX603222
33	14.4	84.7	657	1	AV896818	AV896818 AV896818
34	14.4	84.7	660	8	AQ952838	AQ952838 Sheared D
35	14.4	84.7	665	4	BM612076	BM612076 170006871
36	14.4	84.7	666	2	BE306697	BE306697 601104091
37	14.4	84.7	668	5	BM472135	BM472135 BM472135
38	14.4	84.7	670	4	BM621607	BM621607 170006874
39	14.4	84.7	681	2	BF124082	BF124082 601760175
40	14.4	84.7	685	4	BM632505	BM632505 170006875
41	14.4	84.7	690	5	BM015565	BM015565 BM015565
42	14.4	84.7	706	4	BM652288	BM652288 170006873
43	14.4	84.7	709	4	BM013708	BM013708 BM013708
44	14.4	84.7	714	5	BM607359	BM607359 170006870
45	14.4	84.7	714	8	BZ628070	BZ628070 1N57112.B

ALIGNMENTS

RESULT 1
CD811742
LOCUS
DEFINITION
CD811742
BNI0.00106F011207 BNI0 Brassica napus cDNA clone BNI0001106, mRNA
sequence.

ACCESSION
CD811742
VERSION
CD811742.1 GI:32493682

KEYWORDS
EST.
SOURCE
BRASSICA NAPUS (rape)

ORGANISM
Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE
1 (bases 1 to 588)

AUTHORS
Genoplatte.
TITLE
Unpublished (2003)
JOURNAL
Contact: Genoplatte

COMMENT
Genoplatte

93, rue Henri Rochefort 91025 EVRY CEDEX France

Tel: 33 1 69 47 54 00

Fax: 33 1 69 47 54 10

This sequence has been generated in the framework of the french
plant genomics programme 'genoplatte' (<http://www.genoplatte.com>
and <http://genoplatte-info.inbiolegen.fr>).

Location/Qualifiers

source

1..588
/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="Jec Neuf"
/db_xref="taxon:3708"
/clone="BNI0001106"
/issue_type="seed"
/clone_1ib="BNI0"

ORIGIN

Query Match 94.1%, Score 16, DB 6, Length 588;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCCGCGTAGTAACT 17

DB 526 GCCGCGTAGTAACT 541

RESULT 2
CV101208

LOCUS CV101208 235 bp mRNA linear EST 09-SEP-2004
 DEFINITION UMC-pcrtbcl1-006-f01 Blastocyst after cryopreservation pcrtbcl Sus
 ACCESSION CV101208
 VERSION CV101208.1 GI:51583372
 KEYWORDS EST.
 SOURCE Sus scrofa (pig)
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 REFERENCE Woods, R., Whitworth, K., Springer, G.K., Forrester, L.J.,
 Spollen, W.G., Riese, J., Bivens, N., Murphy, C.N., Mathialigan, N.,
 Agca, Y., Green, J.A. and Prather, R.S.
 TITLE Swine Genomics
 JOURNAL Unpublished (2004)
 COMMENT Contact: DNA Core Facility (Swine Project)
 Animal Science - RS Prather
 University of Missouri-Columbia
 616 Medical Sciences Bldg., Columbia, MO 65212, USA
 Tel: (573) 882-0428
 Fax: (573) 884-5552
 Email: porcine@net.missouri.edu
 POLYA=yes.

FEATURES
 source
 1..235
 /organism="Sus scrofa"
 /mol_type="mRNA"
 /db_xref="taxon:9823"
 /dev_stage="Blastocyst after cryopreservation"
 /clone_id="pcrtbcl1"
 /note="Vector: GRTACTACTA. Funding: A grant from the Monsanto Company to the University of Missouri. Genetic Source: Day 6 pig blastocyst stage embryos were flushed from a gilt's reproductive tract in TV Hapes and cultured in NCSU23 (1) for 3 hours prior to vitrification. For vitrification, embryos were cultured in 1M dimethylsulfoxide and 1.3M ethylene glycol in Tissue Culture Medium (TCM) and 20% newborn calf serum for 5 minutes at 39 degrees Celsius. Then embryos were transferred to 3.2M ethylene glycol and 0.6M sucrose in TCM and 20% newborn calf serum for 45 to 60 seconds at 39 degrees Celsius. Capillary action was used to load the embryos in open pulled straws and the straws were plunged into liquid nitrogen. At thawing embryos were placed in 0.2M sucrose in TCM and 20% newborn calf serum for 1 minute, then 0.1M sucrose in TCM and 20% newborn calf serum for 5 minutes. Embryos were then rinsed and cultured in NCSU23 for 12 or 24 hours. Library construction (PCR protocol): The amount of mRNA that was recovered from embryos was quite limiting and was not sufficient for library production with a standard protocol. Therefore, PCR-based protocol was utilized for producing libraries. The blastocysts were then rinsed with 50 ul polyvinylalcohol-PBS in deep treated medium three times and snap frozen with 1 ul super-RNasin in a 0.5ml RNase-free tube. Hybrid messenger affinity paper was used to isolate the mRNA (2). The mRNA was reverse transcribed with a NotI-tag-dT18 oligonucleotide and a SMART oligonucleotide (Clontech) modified to contain a SalI site to generate full-length cDNA with a sequence complementary to the SMART oligonucleotide. Sequences within the SMART and dT oligonucleotides were used as primers to amplify the cDNAs by PCR with pfu turbo polymerase (Stratagene). The resulting PCR products were purified, digested with NotI and SalI and size fractionated by using a Chroma Spin-400 followed by a Chroma Spin-1000 column (Clontech). Purified cDNA from each PCR reaction was ligated into the pCMV-SPORT6 vector. Preliminary library characterization: Randomly chosen clones from each library were analyzed by restriction digestion to determine average insert size (96 clones) and by sequencing (~3 96-well plates) to confirm library quality (e.g. the presence of short polyA tails,

ORIGIN

Query Match 90.6%; Score 15.4; DB 7; Length 235;
 Best Local Similarity 94.1%; Pred. No. 3.7e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCCGGTAGTACTT 17
 |||||
 Db 146 AGCCCGGTATGTACTT 162

genomic DNA contamination (must be <1%), ribosomal RNA clones (must be <1%), etc.] and to provide a sequence database representing the predominant clones in each library. The clones were sequenced at the University of Missouri-Columbia DNA Core Facility. Bioinformatics work was performed by GK Springer's bioinformatics group (MG Spollen, J Riese, A Gullin, AA Khambati, RV Patel, CM Topinka, SB Bhuiyan) in Computer Science and Health Management and Informatics Departments at the University of Missouri-Columbia. Clone Requests: Requests for clones should be made to the Director of the University of Missouri DNA Core Facility at: porcine@net.missouri.edu. Citations: 1. Im G-S, Lai I, Liu Z, Hao Y and Prather RS. In vitro development of preimplantation porcine nuclear transfer embryos cultured in different media and gas atmospheres. Theriogenology 61: 1125-1135, 2004. 2. Anderson JE, Mattern RL, Abeydeera LR, Day BN and Prather RS. Degradation of maternal Cdc25c during the maternal to zygotic transition is dependent upon embryonic transcription. Molecular Reproduction & Development 60: 181-188, 2001.
 TAG, TISSUE-Blastocyst after cryopreservation
 TAG_SEQ=GTGTAACCTA"

RESULT 3
 BMS39228/c 411 bp mRNA linear EST 20-FEB-2002
 LOCUS hb06f02.g1 Canis CDNA from testes cells Canis familiaris CDNA
 DEFINITION clone hb06f02 5', mRNA sequence.
 ACCESSION BMS39228
 VERSION BMS39228.1 GI:18820982
 KEYWORDS EST.
 SOURCE Canis familiaris (dog)
 ORGANISM Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 REFERENCE O'Shaughnessy, A.L., McCombie, W.R., Baker, J.P., Balija, V.,
 Cummins, D., Dedhia, N.N., de la Bastide, M., Katzenberger, F.,
 King, L., Kirchoff, K.A., Miller, B., Muller, S., Nascimento, L.U.,
 Palmer, L., Santos, L., Shah, R.S., Spiegel, L.A., Zutavern, T.,
 Preston, R. and Hannon, G.J.
 Canis familiaris (dog) (2002)
 EXPRESSED sequence tags from Canis familiaris (dog) (2002)
 JOURNAL Unpublished (2002)
 COMMENT Contact: W. Richard McCombie
 Lita Annenberg Hazen Genome Sequencing Center
 Cold Spring Harbor Laboratory
 PO Box 100 Cold Spring Harbor, NY 11724, USA
 Tel: 516 367 8884
 Fax: 516 367 8874
 Email: mcombie@cshl.org
 plate: hb06 row: F column: 02
 Seq primer: -21M3univrev
 High quality sequence stop: 411.

FEATURES

source
 1..411
 /organism="Canis familiaris"
 /mol_type="mRNA"
 /db_xref="taxon:9615"
 /clone="hb06f02"
 /clone_type="testes"
 /tissue="testes"
 /clone_id="Canis CDNA from testes cells"
 /note="Vector: lambda Zap II; The library was produced by

ORIGIN

Greg Hannon and Raymond Preston (Cold Spring Harbor Laboratory). This library is oligo(dT) primed using stragene zap cDNA synthesis kit. It was made from dog testes. Please contact Greg Hannon (hannon@cshl.org) with any library related inquiries."

Query Match 90.6%; Score 15.4; DB 4; Length 411;
Best Local Similarity 94.1%; Pred. No. 3.8e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACT 17
|||
22 AGCCTCGTAGTAACT 6

Db

RESULT 4 528 bp mRNA linear EST 25-SEP-2003
CA217548/c LOCUS
DEFINITION SCRUD1132D12.b ADI Saccharum officinarum cDNA clone SCRUD1132D12
3', mRNA sequence.
ACCESSION CA217548
VERSION CA217548.1 GI:35267956
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum complex.

REFERENCE 1 (bases 1 to 528)
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE The libraries that made SUCST
JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
Plate: 132 row: D column: 12
Seq primer: SP6 Promoter primer.
Location/Qualifiers

FEATURES
source
1. 528
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCRUD1132D12"
/lab_host="DH10B"
/clone_1lb="AD1"
/note="Organ: seedlings inoculated with Gluconacetobacter diazotrophicans; Vector: pSPORT1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from seedlings inoculated with Gluconacetobacter diazotrophicans). cDNA was prepared from polyA+ mRNA using Superscript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at http://succec.lad.ic.unicamp.br/public"

ORIGIN

Query Match 90.6%; Score 15.4; DB 6; Length 528;
Best Local Similarity 94.1%; Pred. No. 3.8e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACT 17
|||||

Db 330 AGCCGCGTAGTAACT 314

RESULT 5 533 bp mRNA linear EST 21-SEP-2004
CV224495/c LOCUS
DEFINITION CS hyp 19d10 M3Reverse Blue crab hypodermis, normalized
Callinectes sapidus cDNA clone CS hyp 19d10 5', similar to
sp|P81385|CUB_HOMAM Cuticle protein AMP1B (HA-AMP1B). Score = 43.9
bits (102), Expect = 0.002, mRNA sequence.
CV224495
CV224495.1 GI:52371724
EST.
VERSION CV224495.1 GI:52371724
KEYWORDS EST.
SOURCE Callinectes sapidus (blue crab)
ORGANISM Callinectes sapidus (blue crab)
Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
Eubrachyura; Portunoidae; Portunidae; Callinectes.

REFERENCE 1 (bases 1 to 533)
Shafer,T.H., Coblentz,F.R. and Towle,D.W.
TITLE Expressed sequence tags from normalized cDNA libraries prepared from gill and hypodermis tissues of the blue crab, Callinectes sapidus
JOURNAL Unpublished (2004)
COMMENT Contact: Thomas H. Shafer
Department of Biological Sciences
University of North Carolina Wilmington
601 S. College Rd, Wilmington, NC 28403, USA
Tel: 910-962-7275
Fax: 910-962-4066
Email: shafer@uncw.edu
Plate: 19 row: d column: 11
Seq primer: M13 Reverse
High quality sequence stop: 176.
Location/Qualifiers

FEATURES
source
1. 533
/organism="Callinectes sapidus"
/mol_type="mRNA"
/db_xref="taxon:6763"
/clone="CS_hyp_19d10"
/issue_type="Pooled hypodermal epithelium from the mid-dorsal region and arthroal membrane of premolt (stage D2) and 3-hour postmolt crabs"
/dev_stage="Adult"
/clone_1lb="Blue crab hypodermis, normalized"
/note="Vector: pCMV Sport 6.1; Total RNA samples were prepared individually from each tissue, checked for quality, and then pooled for construction and normalization of a cDNA library by Invitrogen. Plasmids were isolated and inserts sequenced from their 5'-ends by the Blue Crab Molecular Genetics Laboratory at the University of North Carolina Wilmington. Traces were trimmed, compared (BLASTx) to NCBI non-redundant protein database as of 19 July 2004, and processed for submission to dbEST by trace2dbEST software (Parkinson, Anthony and Blaxter, unpublished software)."

ORIGIN

Query Match 90.6%; Score 15.4; DB 7; Length 533;
Best Local Similarity 94.1%; Pred. No. 3.8e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACT 17
|||||

Db 469 AGCCACGAGTAGTAACT 453

RESULT 6 667 bp mRNA linear EST 21-SEP-2004
CV224494/c LOCUS
DEFINITION CS hyp 19d10 M3Reverse Blue crab hypodermis, normalized
Callinectes sapidus cDNA clone CS hyp 19d10 5', similar to
sp|P81385|CUB_HOMAM Cuticle protein AMP1B (HA-AMP1B). Score = 36.2
bits (102), Expect = 0.002, mRNA sequence.

QY 1 AGCCGCGTAGTAACT 17
|||||

bites (82), Expect = 0.58, mRNA sequence.
 CV224494
 CV224494.1 GI:52371722
 EST.
 Callinectes sapidus (blue crab)
 SOURCE
 Callinectes sapidus
 Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
 Eubrachyura; Portunoidae; Portunidae; Callinectes.
 REFERENCE
 1 (bases 1 to 667)
 Shafer,T.H., Coblenz,F.B. and Towle,D.W.
 Expressed sequence tags from normalized cDNA libraries prepared
 from gill and hypodermis tissues of the blue crab, Callinectes
 sapidus
 JOURNAL
 Unpublished (2004)
 COMMENT
 Contact: Thomas H. Shafer
 Department of Biological Sciences
 University of North Carolina Wilmington
 601 S. College Rd. Wilmington, NC 28403, USA
 Tel: 910-962-7275
 Fax: 910-962-4066
 Email: shafer@uncw.edu
 Plate: 19 row: d column: 10
 Seq primer: M13 Reverse
 High quality sequence stop: 494.
 Location/Qualifiers
 1..667
 /organism="Callinectes sapidus"
 /mol_type="mRNA"
 /db_xref="taxon:6763"
 /clone="CS_hyp_19d10"
 /tissue_type="Pooled hypodermal epithelium from the
 mid-dorsal region and arthropodal membrane of premolt
 (stage D2) and 3-hour postmolt crabs"
 /dev_stage="Adult"
 /clone_lib="Blue crab hypodermis, normalized"
 /note="Vector: PCWY Sport 6.1; Total RNA samples were
 prepared individually, from each tissue, checked for
 quality, and then pooled for construction and
 normalization of a cDNA library by Invitrogen. Plasmids
 were isolated and inserts sequenced from their 5'-ends by
 the Blue Crab Molecular Genetics Laboratory at the
 University of North Carolina Wilmington. Traces were
 trimmed, compared (BLASTX) to NCBI non-redundant protein
 database as of 19 July 2004, and processed for submission
 to dbEST by trace2dbEST software (Parkinson, Anthony and
 Blaxter, unpublished software)."
 ORIGIN
 Query Match 90.6%; Score 15.4; DB 7; Length 667;
 Best Local Similarity 94.1%; Pred. No. 3.9e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AGCCGGGTAGTAACT 17
 |||||
 Db 468 AGCCAGCTAGTAACT 452
 RESULT 7
 BQ294905/c 700 bp mRNA linear EST 16-MAY-2002
 LOCUS BQ294905
 DEFINITION WHE2855.G07 M13S wheat unexpressed root tip cDNA library Triticum
 aestivum cDNA clone WHE2855_G07_M13, mRNA sequence.
 ACCESSION BQ294905
 VERSION BQ294905.1 GI:20810125
 KEYWORDS EST.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooidae; Triticaceae; Triticum.
 1 (bases 1 to 700)
 Anderson,O.D., Chao,S., Chin,A., Close,T.J., Crossman,C.,

Guatfson,P., Lazo,G.R., Pham,J., Rauech,C.J., Ross,K., Wilson,C.
 and Moo,J.
 The structure and function of the expressed portion of the wheat
 genomes - Unexpressed root tip cDNA library
 Unpublished (2002)
 JOURNAL
 CONTACT: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105959773
 Fax: 5105959818
 Email: oanderson@pw.usda.gov
 Sequences have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: SK primer
 Location/Qualifiers
 1..700
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE2855_G07_M13"
 /tissue_type="Root tip at 1.0 to 1.5 mm stage"
 /dev_stage="Four-day old seedling"
 /lab_host="E. coli SOLR"
 /clone_lib="Wheat unexpressed root tip cDNA library"
 /note="Vector: Lambda Uni-ZAP XR, excised phagmid;
 Site_1: EcoRI; Site_2: XhoI; Plants were grown under
 hydroponic conditions for four days. Root tips were
 excised and snap frozen (Ross and Guatfson) and total RNA
 was prepared at University of Missouri, Columbia. Poly(A)
 RNA was purified, a cDNA library was made, and the cDNA
 clones were in vivo excised to give phluescript
 SK(-) phgemids in the T7 Close lab (Chin and Close) at the
 University of California, Riverside. Plasmid DNA
 preparations and DNA sequencing were performed in the OD
 Anderson lab (all other authors)."
 ORIGIN
 Query Match 90.6%; Score 15.4; DB 5; Length 700;
 Best Local Similarity 94.1%; Pred. No. 3.9e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AGCCGGGTAGTAACT 17
 |||||
 Db 24 AGCCGGGTAGTAACT 8
 RESULT 8
 CL653433 768 bp DNA linear GSS 09-JUL-2004
 LOCUS PR10118c_B08 - PR10118c_B21 (768) Mixed stage fosmid library of P.
 pacificus var. California Pristionchus pacificus genomic, genomic
 survey sequence.
 ACCESSION CL653433
 VERSION CL653433.1 GI:50132271
 KEYWORDS GSS.
 SOURCE Pristionchus pacificus
 ORGANISM Pristionchus pacificus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
 Neodiplogasterida; Pristionchus.
 1 (bases 1 to 768)
 Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
 AppADB: an AcedB database for the nematode satellicite organism
 Pristionchus pacificus
 Nucleic Acids Res. 32 (1), D421-D422 (2004)
 JOURNAL
 CONTACT: Sommer RJ
 Evolutionary Biology
 Max-Planck-Institute for Developmental Biology
 Spemannstr. 37-39, Tuebingen D-72076, Germany
 Tel: 00497071601371
 Fax: 00497071601498
 Email: ralf.sommer@tuebingen.mpg.de

This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid end.

FEATURES

source
Location/Qualifiers

1..768
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pBplfos-5 Fosmid vector"

ORIGIN

Query Match 90.6%; Score 15.4; DB 9; Length 768;
Best Local Similarity 94.1%; Pred. No. 3.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACT 17
|||||
Db 542 AGCCGCGTAGTAACT 526

RESULT 9 509 bp mRNA linear EST 03-JUN-2003
LOCUS CD445870
DEFINITION EL0170204F06.b Endosperm_4 Zea mays cDNA, mRNA sequence.
ACCESSION CD445870
VERSION CD445870.1 GI:31361513
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
1 (bases 1 to 509)
Lai J., Dey, N., Kim, C.S., Bharti, A.K., Rudd, S., Mayer, K.F.,
Larting, B., Becraft, P. and Messing, J.
Characterization of the maize endosperm transcriptome and its
comparison to the rice genome
Genome Res. 14 (10), 1932-1937 (2004)

JOURNAL Contact: Lai, Jinheng
Dr. Joachim Messing's lab
Makeman Institute, Rutgers University
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Tel: 732-445-3801
Fax: 732-445-5735
Email: jlai@waksman.rutgers.edu
Seq primer: T3.

FEATURES
source
Location/Qualifiers

1..509
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="W22"
/db_xref="taxon:4577"
/tissue_type="Endosperm of 7-23DAP"
/clone_lib="Endosperm_4"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

ORIGIN

Query Match 88.2%; Score 15; DB 6; Length 509;
Best Local Similarity 100.0%; Pred. No. 6.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACT 15
|||||
Db 74 AGCCGCGTAGTAACT 88

RESULT 10
CNS06U7P

LOCUS CNS06U7P 724 bp DNA linear GSS 06-JUN-2001
DEFINITION T3 end of clone AX0AA005G05 of library AX0AA from strain CBS 7064
of Pichia farinosa, genomic survey sequence.

ACCESSION AL415499
VERSION AL415499.1 GI:12194769
KEYWORDS GSS.
SOURCE Pichia farinosa
ORGANISM Pichia farinosa

REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.
Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
1 (bases 1 to 724)
Bolotin-Fukuhara, M., Bon, E., Brotier, P., Casaregola, S.,
de Montigny, J., Dujon, B., Durans, P., Lepingle, A., Llorente, B.,
Malpertuy, A., Nevegilise, C., Ozier-Kalogeropoulos, O., Pottier, S.,
Saurin, W., Tekala, F., Toffano-Nicche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissenbach, J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEMS Lett. 487 (1), 3-12 (2000)

JOURNAL MEDLINE
PUBMED 20584711
TITLE 2 (bases 1 to 724)
REFERENCE de Montigny, J., Spelmer, C., Souciet, J., Tekala, F., Dujon, B.,
Wincker, P., Artiguenave, F. and Potier, S.
Genomic exploration of the hemiascomycetous yeasts: 15. Pichia
sorbitophila
FEMS Lett. 487 (1), 87-90 (2000)

JOURNAL MEDLINE
PUBMED 20584725
TITLE 3 (bases 1 to 724)
REFERENCE Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source
Location/Qualifiers

1..724
/organism="Pichia farinosa"
/mol_type="genomic DNA"
/strain="CBS 7064"
/db_xref="taxon:4920"
/clone_lib="AX0AA005G05"
/clone_lib="AX0AA"
/note="end : T3"

ORIGIN

Query Match 88.2%; Score 15; DB 9; Length 724;
Best Local Similarity 100.0%; Pred. No. 6.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCGCGTAGTAACT 17
|||||
Db 207 CCGCGTAGTAACT 221

RESULT 11 821 bp DNA linear GSS 17-MAR-2003
LOCUS BZ797610/
DEFINITION PUFCT697D_ZM_0.6_1.0_KB Zea mays genomic clone ZMBFaz291K17,
genomic survey sequence.
ACCESSION BZ797610
VERSION BZ797610.1 GI:29000148

```

KEYWORDS      GSS.
SOURCE         Zea mays
ORGANISM       Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
                clade; Panicoideae; Andropogoneae; Zea.
REFERENCE      1 (bases 1 to 821)
                Whiteaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,
                Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
                Bennettzen,J.
                Maize Genomics Consortium
                Unpublished (2003)
COMMENT        Other GSSs: PUCY69TB
                Contact: Cathy Whiteaw
                TIGR
                9712 Medical Center Drive, Rockville, MD 20850, USA
                Tel: 301-838-5843
                Fax: 301-838-0208
                Email: whiteaw@tigr.org
                Seq primer: TF
                Class: sheared ends.
FEATURES       Location/Qualifiers
                source          1..821
                                /organism="Zea mays"
                                /mol_type="genomic DNA"
                                /strain="B73"
                                /db_xref="taxon:4577"
                                /clone="ZMBRta291K17"
                                /clone_1lb="ZM 0.6_1.0 KB"
                                /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
                                Cor selected genomic DNA library"
ORIGIN
Query Match      88.2%; Score 15; DB 8; Length 821;
Best Local Similarity 100.0%; Pred. No. 6.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCGCGTAGGTAC 15
        |||||||
        323 AGCCGCGTAGGTAC 309

RESULT 12      927 bp      DNA      linear      GSS 19-AUG-2003
CG063514      PUKAR59TD_ZM_0.6_1.0_KB Zea mays genomic clone ZMBRta0776122,
LOCUS          genomic survey sequence.
DEFINITION     CG063514
ACCESSION      CG063514
VERSION        CG063514.1
KEYWORDS       GSS.
SOURCE         Zea mays
ORGANISM       Zea mays
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
                clade; Panicoideae; Andropogoneae; Zea.
                1 (bases 1 to 927)
                Whiteaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,
                Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
                Bennettzen,J.
                Maize Genomics Consortium
                Unpublished (2003)
COMMENT        Other GSSs: PUKAR59TB
                Contact: Cathy Whiteaw
                TIGR
                9712 Medical Center Drive, Rockville, MD 20850, USA
                Tel: 301-838-5843
                Fax: 301-838-0208
                Email: whiteaw@tigr.org
                Seq primer: TF
                Class: sheared ends.
FEATURES       Location/Qualifiers
                source          1..927
                                /organism="Zea mays"

```

```

                                /mol_type="genomic DNA"
                                /strain="B73"
                                /db_xref="taxon:4577"
                                /clone="ZMBRta0776122"
                                /clone_1lb="ZM 0.6_1.0 KB"
                                /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
                                Cor selected genomic DNA library"
ORIGIN
Query Match      88.2%; Score 15; DB 9; Length 927;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCGCGTAGGTAC 15
        |||||||
        737 AGCCGCGTAGGTAC 723

RESULT 13      1107 bp      mRNA      linear      EST 21-DEC-2000
BF672870      602152718r1 NIH_MGC_81 Homo sapiens CDNA clone IMAGE:4293820 5',
LOCUS          mRNA sequence.
DEFINITION     BF672870.1 GI:11946765
ACCESSION      BF672870
VERSION        BF672870.1
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
                1 (bases 1 to 1107)
                NIH-MGC http://mgs.nci.nih.gov/.
                National Institutes of Health, Mammalian Gene Collection (MGC)
                Unpublished (1999)
                Contact: Robert Strausberg, Ph.D.
                Email: cga@bbs-rt@mail.nih.gov
                Tissue Procurement: CLONETECH Laboratories, Inc.
                CDNA Library Preparation: CLONETECH Laboratories, Inc.
                CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LNL at:
                http://image.llnl.gov
                Plate: LNCM1142 row: c column: 05
                High quality sequence start: 33
                High quality sequence stop: 49.
FEATURES       Location/Qualifiers
                source          1..1107
                                /organism="Homo sapiens"
                                /mol_type="mRNA"
                                /db_xref="taxon:9606"
                                /clone="IMAGE:4293820"
                                /lab_host="DH10B (T1 phage-resistant)"
                                /clone_1lb="NIH_MGC_81"
                                /note="Organ: muscle (skeletal); Vector: pDNR-LIB
                                (Clontech); Site 1: SfiI (ggcgctcgcc); Site 2: SfiI
                                (ggcgctatggcc); 5' and 3' adaptors were used in cloning
                                as follows: 5'-adaptor sequence: 5'-CACGGCATTATGGC-3'
                                and 3'-adaptor sequence:
                                5'-ATTCTAGAGCGAGCGCGGCACATG-dt(30)BN-3' (where B = A,
                                C, or G and N = A, C, G, or T). Average insert size
                                1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
                                inserts by PCR. This library was enriched for full-length
                                clones and was constructed by Clontech Laboratories (Palo
                                Alto, CA)."
ORIGIN
Query Match      88.2%; Score 15; DB 2; Length 1107;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCGCGTAGGTAC 15
        |||||||

```

Db 814 AGCCGCGTAGTAAC 828

RESULT 14
 CK914339 398 bp mRNA linear EST 22-APR-2004
 DEFINITION p1fmgcf_000943 Normalized Magnaporthe grisea cDNA p1fmgcf_000943, mRNA sequence.
 CK914339
 MAGNAPORTHE GRISEA
 VERSION CK914339
 KEYWORDS CK914339.1 GI:45373044
 SOURCE EST.
 ORGANISM Magnaporthe grisea (anamorph: Pyricularia grisea)
 Magnaporthe grisea
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariomycetes; Incertae sedis; Magnaporthaceae; Magnaporthe.
 1 (bases 1 to 398)
 Dong, H., Peng, X., Chen, B., Li, Y. and Li, D.
 Large-scale identification of ESTs from Magnaporthe grisea by
 normalized cDNA library sequencing (2004b)
 Unpublished (2004)
 CONTACT: Debao Li
 Bioinformatics and Gene Network Research Group
 Zhejiang University, China Agricultural University, Guangxi
 University
 Kaixuan Road 268#, Hangzhou, Zhejiang, 310029, China
 Tel: 0086-571-86961525
 Fax: 0086-571-86971183
 Email: webmaster@estarray.org, URL: http://www.estarray.org
 Seq primer: M13 reverse primer.
 Location/Qualifiers
 1..398
 /organism="Magnaporthe grisea"
 /mol_type="mRNA"
 /db_xref="taxon:148305"
 /clone="p1fmgcf_000943"
 /issue_type="MycoCellium, conidium, germinating conidium,
 swelling appressorium, mature appressorium, penetration
 peg"
 /dev_stage="MycoCellium, conidium, germinating conidium,
 swelling appressorium, mature appressorium, penetration
 peg"
 /clone_lib="Normalized Magnaporthe grisea cDNA pRS423
 library"
 /note="Vector: pRS423"

ORIGIN

Query Match 84.7%; Score 14.4; DB 7; Length 398;
 Best Local Similarity 93.8%; Pred. No. 1.5e+03;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCGCGTAGTAACCT 17
 ||||||||||||
 Db 139 GCGCGTAGTAACCT 154

RESULT 15
 AJ828384 419 bp mRNA linear EST 07-SEP-2004
 LOCUS AJ828384/c
 DEFINITION AJ828384 HaCoCR Helianthus annuus cDNA clone HaCoCR005B01, mRNA
 sequence.
 ACCESSION AJ828384
 VERSION AJ828384
 KEYWORDS AJ828384.1 GI:51914389
 SOURCE EST.
 ORGANISM Helianthus annuus (common sunflower)
 Helianthus annuus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; campanulids; Asterales; Asteraceae; Asteroideae;
 Heliantheae; Helianthus.
 1 (bases 1 to 419)
 Ben, C., Hwezel, T., Tamborindeguy, C. and Gentschitel, L.
 Comparative analysis of developing sunflower seed cDNA libraries
 Unpublished (2004)

COMMENT Contact: Ben C
 Laboratoire BAP
 Inpt - Ensat, Pole de Biotechnologie Vegetale
 18 Chemin de Borde Rouge, Auzerville, Castanet Tolosan 31326,
 FRANCE.

FEATURES
 source Location/Qualifiers
 1..419
 /organism="Helianthus annuus"
 /mol_type="mRNA"
 /db_xref="taxon:4232"
 /clone="HaCoCR005B01"
 /dev_stage="cotyledonary embryo"
 /clone_lib="HaCoCR"
 /note="Helianthus annuus cotyledonary embryo; indred line:
 Ha300b"

ORIGIN

Query Match 84.7%; Score 14.4; DB 1; Length 419;
 Best Local Similarity 93.8%; Pred. No. 1.5e+03;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCGCGTAGTAACC 16
 ||||||||||||
 Db 196 AGCCGCGTAGTAACC 181

Search completed: November 7, 2005, 09:18:56
 Job time: 1708.72 secs

2.5 Egg Bunk (uspiq)